

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 10:38:48 ; Search time 433.634 Seconds
(without alignments)
15657.843 Million cell updates/sec

Title: US-09-762-767a-1
Perfect score: 3015
Sequence: 1 ccctatgctgattggcgg.....aaaaaaaaaaaaaaaa 3015

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues 4370478
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3015	100.0	3015	21	Human soluble guan
2	2950.8	97.9	4742	23	Human prostate exp
3	2950.8	97.9	4742	23	Human prostate exp
4	2950.8	97.9	4742	23	Human prostate exp
5	2950.8	97.9	4742	23	Human prostate exp
6	2855.4	94.7	3004	24	Prostate cancer-as
7	2855.4	94.7	3004	24	Prostate cancer-as
8	2368	78.5	2715	21	Human cyclic nucle
9	745.8	24.7	885	22	Human neuroblastom

10	548	18.2	2954	24	AAS94827	Human DNA sequence
11	482.2	16.0	761	22	ABL23886	Human breast cance
12	315.6	10.5	2357	23	ABL02451	Drosophila melanog
13	314.8	10.4	5259	23	ABL02450	Drosophila melanog
14	308.4	10.2	649	22	ABL15033	Human breast cance
15	260.8	8.7	3094	23	ABL04726	Drosophila melanog
16	204.4	6.8	751	22	ABL19789	Human neuroblastom
17	192.6	6.4	2443	21	AAZ88938	Human soluble guan
18	192.6	6.4	3196	22	AAH98392	Human EST-derived
19	176.4	5.9	2335	24	ABK63662	Rat sequence diffe
20	155.6	5.2	2850	23	ABL02865	Drosophila melanog
21	152.4	5.1	5424	24	ABL32854	Human immune syste
22	149.4	5.0	3821	22	ABA09027	Human ANP-A recept
23	148.6	4.9	3519	23	ABL30435	Drosophila melanog
24	148.6	4.9	6756	23	ABL30434	Drosophila melanog
25	147	4.9	2212	23	ABL07087	Drosophila melanog
26	146	4.8	3697	12	AAQ10324	Human Natriuretic
27	144.4	4.8	4081	24	ABR83833	Human cDNA diffe
28	139.8	4.6	3231	23	ABL05735	Drosophila melanog
29	139.8	4.6	6339	23	ABL05734	Drosophila melanog
30	139.6	4.6	2928	23	ABL15689	Drosophila melanog
31	135	4.5	3286	23	ABL28779	Drosophila melanog
32	134	4.5	8895	23	ABL28778	Drosophila melanog
33	134	4.4	2766	23	ABL29938	Drosophila melanog
34	128.8	4.3	3762	22	AAF89876	Nucleotide sequenc
35	127.4	4.3	5565	23	ABL29715	Drosophila melanog
36	127.4	4.2	5424	24	ABL32855	Drosophila melanog
37	125.6	4.2	3784	14	AAQ47356	Human immune syste
38	124.6	4.1	3499	22	ABF89878	GC-C DNA. Rattus
39	121.4	4.0	2664	23	AAS67254	Nucleotide sequenc
40	121.4	4.0	2664	23	AAS83974	DNA encoding novel
41	121.4	4.0	2664	23	AAS89581	DNA encoding novel
42	121.4	4.0	3786	18	AAH89300	Human ST receptor
43	121.4	4.0	3786	19	AAH97329	Human ST receptor
44	121.4	4.0	3787	24	ABL57068	Human ST receptor
45	121.4	4.0	3787	24	AAS16116	DNA encoding human

ALIGNMENTS

RESULT 1
AAZ88938
ID AAZ88938 standard; DNA; 3015 BP.

AC AAZ88938;

DT 26-MAY-2000 (first entry)

XX Human soluble guanylylcyclase alpha DNA.

XX Human; guanylylcyclase alpha; hsgCalphal; soluble;
KW guanylylcyclase beta; antiarteriosclerotic; vasotropic; hypotensive;
KW gene therapy; arteriosclerosis; restenosis; ischemia; diagnosis;
KW peripheral arterial occlusive disease; arterial hypertension; ds.

XX Homo sapiens.

OS DE19837015-A1.

PN 24-FEB-2000.

PD 14-AUG-1998; 98DE-1037015.

XX 14-AUG-1998; 98DE-1037015.

XX (VASO-) VASOPHARM BIOFTECH GMBH & CO KG.

XX Schmidt H, Zabel U, Poller W;

XX WPI; 2000-184044/17.

XX P-PSDB; AAY51607.

QY 1861 GAAGAAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCT 1920
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1861 GAAGAAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCT 1920
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1921 GTGGCAAGGCAAGTTGCAAGCAAGCAAGTTCAAGTATGTCACCATGCTCTCTCAGA 1980
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1921 GTGGCAAGGCAAGTTGCAAGCAAGCAAGTTCAAGTATGTCACCATGCTCTCTCAGA 1980
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1981 CATCGTTGGTTCACTGCCATCTGCTCCAGTGTCTACCGCTGCAGGTCATCACCATGCT 2040
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2041 CAATGCACGTACACTGCTCGACAGCAGTGTGGAGAGCTGATGTCACAGGTGA 2100
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2041 CAATGCACGTACACTGCTCGACAGCAGTGTGGAGAGCTGATGTCACAGGTGA 2100
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2101 GACCATGCGCATGCTTATGTTAGCTGGGGATTACAAAGAGAGTGTACTCATGTC 2160
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2101 GACCATGCGCATGCTTATGTTAGCTGGGGATTACAAAGAGAGTGTACTCATGTC 2160
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2161 TGTTCCAGATAGCGCTGATGGCCCTGAAGATGATGAGCTCTCTGATGAAGTTATGCTCC 2220
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2161 TGTTCCAGATAGCGCTGATGGCCCTGAAGATGATGAGCTCTCTGATGAAGTTATGCTCC 2220
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2221 CCATGGAGACCTATCAAGATGCGAATTTGGAGTGCACCTCTGGATCAGTTTTCGCGGT 2280
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2221 CCATGGAGACCTATCAAGATGCGAATTTGGAGTGCACCTCTGGATCAGTTTTCGCGGT 2280
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2281 CGTTGGAGTTAAATGCCCCCTTACTGCTTTTGGAAACATGTCACCTGCGCTAACAA 2340
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2281 CGTTGGAGTTAAATGCCCCCTTACTGCTTTTGGAAACATGTCACCTGCGCTAACAA 2340
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2341 ATTTGAGTCCCTGAGTACCAAGAAAATCAATGTCAGCCCAACCACTTACAGATTACT 2400
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2341 ATTTGAGTCCCTGAGTACCAAGAAAATCAATGTCAGCCCAACCACTTACAGATTACT 2400
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2401 CAAGACACTGCTGCTGTTGCTGTTTACCCCTCGATCAAGGAGGAACTTCCACCAACTT 2460
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2401 CAAGACACTGCTGCTGTTGCTGTTTACCCCTCGATCAAGGAGGAACTTCCACCAACTT 2460
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2461 CCTAGTGAATATCCCGGAATCTGCAATTTCTGATGCTTACCAACAGGAAACCACTC 2520
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2461 CCTAGTGAATATCCCGGAATCTGCAATTTCTGATGCTTACCAACAGGAAACCACTC 2520
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2521 AAACCCATGCTTCCAAAAGAAAGATGTGGAAGATGCAATGCCAATTTTATAGGCAAGC 2580
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2521 AAACCCATGCTTCCAAAAGAAAGATGTGGAAGATGCAATGCCAATTTTATAGGCAAGC 2580
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2581 ATCAGGAATAGATTAGCAACCTATATACCTATTTATAGTCTTTGGGTTGACTCATTTG 2640
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2581 ATCAGGAATAGATTAGCAACCTATATACCTATTTATAGTCTTTGGGTTGACTCATTTG 2640
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2641 AAGATGTGTAGAGCTCTGAAGACACTTTAGGATGTAGATGGCTTAACAAGCAGTATTA 2700
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2641 AAGATGTGTAGAGCTCTGAAGACACTTTAGGATGTAGATGGCTTAACAAGCAGTATTA 2700
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2701 AAATTTAGGAGCCCAAGTACAACTTTCTCTCTGTTTAACTGACAAAATGACTCATTT 2760
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2701 AAATTTAGGAGCCCAAGTACAACTTTCTCTCTGTTTAACTGACAAAATGACTCATTT 2760
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2761 CAGTACTTCAGCTCTTCAGAAAACCAACCTTAAAGAGCTTAAAGAGCTTTTGGGAGTA 2820
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2761 CAGTACTTCAGCTCTTCAGAAAACCAACCTTAAAGAGCTTAAAGAGCTTTTGGGAGTA 2820
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2821 TTCTTATTATATACCAAGCACTTACTTACCTGTACTCAAAATTCAGCACCTTGTACATATA 2880
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2821 TTCTTATTATATACCAAGCACTTACTTACCTGTACTCAAAATTCAGCACCTTGTACATATA 2880
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2881 TCAGATAATTTAGTCAATTTGACAACTGTATGGAGTCCACCTGCAATCTCATATCTCTG 2940
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2881 TCAGATAATTTAGTCAATTTGACAACTGTATGGAGTCCACCTGCAATCTCATATCTCTG 2940
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 2941 GGAATGCCATGTTTATTAAAGTGTCTTTGTGATAGTGTCTGTCACAAAAA 3000
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2941 GGAATGCCATGTTTATTAAAGTGTCTTTGTGATAGTGTCTGTCACAAAAA 3000
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 3001 AAAAAA 3015
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 3001 AAAAAA 3015
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RESULT 2
ABV21348
ID ABV21348 standard; cDNA; 4742 BP.
XX AC ABV21348;
XX DT 13-SEP-2002 (first entry)
XX Human prostate expression marker cDNA 21339.
DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
KW Homo sapiens.
OS WO200160860-A2.
PN 23-AUG-2001.
PD 20-FEB-2001; 2001WO-US05171.
XX PF 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
DR Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX Claim 1; Page 3547; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX Sequence 4742 BP; 1450 A; 933 C; 998 G; 1355 T; 6 other;
Query Match 97.9%; Score 2950.8; DB 23; Length 4742;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2978; Conservative 0; Mismatches 27; Indels 1; Gaps 1;
QY 6 ATGGCGATTGGCGGCTGCAGAGACAGGACATCAGTCCCTGCTCTGAGCCTAG 65
|||||

Db 6 ATGGCGATTTGGGGCTGACAGACACAGGACTCAGTTCCCTGCCCTAGTCTGAGCCTAG 65
QY 66 TGGTGGGACTCAGCTCAGAGTCAGTTTTCCAGAAGCAGGTTTCAGTGCAGAGTTTTCCT 125
Db 66 TGGTGGGACTCAGCTCAGAGTCAGTTTTCCAGAAGCAGGTTTCAGTGCAGAGTTTTCCT 125
QY 126 ACATTTTCTGCGCTAGACAGCAGGAGCCTTGGAAACAGACCCAGCGGAGGACACCTG 185
Db 126 ACATTTTCTGCGCTAGACAGCAGGAGCCTTGGAAACAGACCCAGCGGAGGAGCAGCTG 185
QY 186 TGGGGAGGAGCGCTTGGAGGAGCTTAGAGACCCAGCGCGGCTGATCTCACCAGTGG 245
Db 186 TGGGGAGGAGCGCTTGGAGGAGCTTAGAGACCCAGCGCGGCTGATCTCACCAGTGG 245
QY 246 CGGATTTGCGAGCGCGCTTGGAGCTGCTAGAGATCCGGAACACAGCCCGAGGTGG 305
Db 246 CGGATTTGCGAGCGCGCTTGGAGCTGCTAGAGATCCGGAACACAGCCCGAGGTGG 305
QY 306 CGAAGCCCAAGACTCGCGCTTGGAGAAAGCTGAGCAGGGGGCCAGCGGCTGCC 365
Db 306 CGAAGCCCAAGACTCGCGCTTGGAGAAAGCTGAGCAGGGGGCCAGCGGCTGCC 365
QY 366 GCGCTGCTGCAACCTGTCGCTGAGCTGCTGACAGTGACATGACATCCCGAGTTACC 425
Db 366 GCGCTGCTGCAACCTGTCGCTGAGCTGCTGACAGTGACATGACATCCCGAGTTACC 424
QY 426 AGTGTCTTGAATGATGAGTGGCTTCTGTTGTGAGTCTCATATAAGAACTACAGCTCAT 485
Db 426 AGTGTCTTGAATGATGAGTGGCTTCTGTTGTGAGTCTCATATAAGAACTACAGCTCAT 485
QY 486 CAGGAGGAGATCCAGCAGGCTTGAAGACACCAACACCATGTTCTGCAGCAAGCTCAAG 545
Db 486 CAGGAGGAGATCCAGCAGGCTTGAAGACACCAACACCATGTTCTGCAGCAAGCTCAAG 544
QY 546 ATCTCAAGATCAGAGAGTGTCTTCTCTTCTGCTGACAGGCTCAAGTTCTTAAG 605
Db 546 ATCTCAAGATCAGAGAGTGTCTTCTCTTCTGCTGACAGGCTCAAGTTCTTAAG 604
QY 606 AGTCTTCAAGAGGAGCAGAGAGTGTCTTCTCTTCTGCTGACAGGCTCAAGTTCTTAAG 665
Db 606 AGTCTTCAAGAGGAGCAGAGAGTGTCTTCTCTTCTGCTGACAGGCTCAAGTTCTTAAG 664
QY 666 AAGACATCTCTGAGAGAACATCAAGAAAGTCTTCTCTCAAGAAACACAGTCGAGGCC 725
Db 666 AAGACATCTCTGAGAGAACATCAAGAAAGTCTTCTCTCAAGAAACACAGTCGAGGCC 724
QY 726 GAGTCTATCTTCACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCAGAGTTTGAAC 785
Db 726 GAGTCTATCTTCACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCAGAGTTTGAAC 784
QY 786 GCGTGAATGTTGACATTCAGAGAACATTTGCAAAAGCAGCAAAATATAAGAAAGCAGGAAT 845
Db 786 GCGTGAATGTTGACATTCAGAGAACATTTGCAAAAGCAGCAAAATATAAGAAAGCAGGAAT 844
QY 846 CTTTGGAAAGAGAGACTTTGAAACAACTTCCAGAGCAGCAGTTGCAAGCAGGAGTTTC 905
Db 846 CTTTGGAAAGAGAGACTTTGAAACAACTTCCAGAGCAGCAGTTGCAAGCAGGAGTTTC 904
QY 906 CAGTGGAGGTTATCAAGAACTCTTGGTGAAGAGGTTTTTAAATATGTTACAGGAAG 965
Db 906 CAGTGGAGGTTATCAAGAACTCTTGGTGAAGAGGTTTTTAAATATGTTACAGGAAG 964
QY 966 ATGAAACATCTTGGGGTGGTGGAGGACCTTAAAGATTTTTTAAACAGCTTCAGTA 1025
Db 966 ATGAAACATCTTGGGGTGGTGGAGGACCTTAAAGATTTTTTAAACAGCTTCAGTA 1024
QY 1026 CCCTTCTGAACAGACAGCAGCTTGCAGAAAGCAGCAAGAAAGGCGGCTTGAAGAGC 1085
Db 1026 CCCTTCTGAACAGACAGCAGCTTGCAGAAAGCAGCAAGAAAGGCGGCTTGAAGAGC 1084
QY 1086 CTTCCATCTTATGCTTGGATTAAGAGGATGATTTTCTACATGTTTACTACTTCTCCCTA 1145
Db 1086 CTTCCATCTTATGCTTGGATTAAGAGGATGATTTTCTACATGTTTACTACTTCTCCCTA 1144

QY 1146 AGAAGAACACCTCCCTGATTTTCCGGGATCATAAAGCAGCTGCTCAGCTATTATATG 1205
Db 1146 AGAAGAACACCTCCCTGATTTTCCGGGATCATAAAGCAGCTGCTCAGCTATTATATG 1204
QY 1206 AAACGGAAGTGGAGTGTGTTAATGCTCCCTGCTTCCATATATGATGTCAGGAGTTG 1265
Db 1206 AAACGGAAGTGGAGTGTGTTAATGCTCCCTGCTTCCATATATGATGTCAGGAGTTG 1264
QY 1266 TGAATCAGCCTACTTGTGTGTTTACTCCTGTTTCCATGAAAGCAGCAAGCCTCCTGTC 1325
Db 1266 TGAATCAGCCTACTTGTGTGTTTACTCCTGTTTCCATGAAAGCAGCAAGCCTCCTGTC 1324
QY 1326 CCAGCAAAACCCAGTCTGCTGCTGATTTCCACATGCTTATTTCTCAAGACATTTCCAT 1385
Db 1326 CCAGCAAAACCCAGTCTGCTGCTGATTTCCACATGCTTATTTCTCAAGACATTTCCAT 1384
QY 1386 TCCATTTTCATTTTGACAAAGATATGACAAATTTGCAATTTGGCAATGGCATCAGAGGC 1445
Db 1386 TCCATTTTCATTTTGACAAAGATATGACAAATTTGCAATTTGGCAATGGCATCAGAGGC 1444
QY 1446 TGATGAACAGGAGAGACTTTCAAGAAAGCTTAATTTTGAAGAAATACTTTGAAATTTCTGA 1505
Db 1446 TGATGAACAGGAGAGACTTTCAAGAAAGCTTAATTTTGAAGAAATACTTTGAAATTTCTGA 1504
QY 1506 CTCCAAAATCAACAGAGCTTTAGCGGGATCATGACTATGTTGAAATATGCACTTTGTTG 1565
Db 1506 CTCCAAAATCAACAGAGCTTTAGCGGGATCATGACTATGTTGAAATATGCACTTTGTTG 1564
QY 1566 TAGCAGTGGAGAGATGGGACAACTCTGTGAGAAATCTTCAAGGGTTATGGACCTCAAG 1625
Db 1566 TAGCAGTGGAGAGATGGGACAACTCTGTGAGAAATCTTCAAGGGTTATGGACCTCAAG 1624
QY 1626 GCCAAATGATCTACATTTGTAATCCAGTGCAATCTGTTTGGGTGCTACCTGCTGTGG 1685
Db 1626 GCCAAATGATCTACATTTGTAATCCAGTGCAATCTGTTTGGGTGCTACCTGCTGTGG 1684
QY 1686 ACAGATTAGAAGATTTTACAGGAGGCTCTACCTCTCAGACATCCCAATTTCAATG 1745
Db 1686 ACAGATTAGAAGATTTTACAGGAGGCTCTACCTCTCAGACATCCCAATTTCAATG 1744
QY 1746 CACTGAGGATGTGGTCTTAATAGGGAACAAAGCAGCTCAAGTGGCTCAAGAA 1805
Db 1746 CACTGAGGATGTGGTCTTAATAGGGAACAAAGCAGCTCAAGTGGCTCAAGAA 1804
QY 1806 GGCTGGGAGCTGAAGGCTTACCTTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAA 1865
Db 1806 GGCTGGGAGCTGAAGGCTTACCTTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAA 1864
QY 1866 AAAGACAGTAGACCTTCTGCTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCTGTGGC 1925
Db 1866 AAAGACAGTAGACCTTCTGCTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCTGTGGC 1924
QY 1926 AAGGGCAAGTTGTCAGCAAGCAAGAACTTCAATGTCACCATGCTTCTCAGACATCG 1985
Db 1926 AAGGGCAAGTTGTCAGCAAGCAAGAACTTCAATGTCACCATGCTTCTCAGACATCG 1984
QY 1986 TTGGGTTCACTGCTGCTGCTCCAGTGTCTCAGCTGCTCAGGTCATCACCCTCTCAATG 2045
Db 1986 TTGGGTTCACTGCTGCTGCTCCAGTGTCTCAGGTCATCACCCTCTCAATG 2044
QY 2046 CACTGTACACTCGCTTTCAGCAGCTGTGGAGAGCTGTTACAAAGGTTGAGAGCA 2105
Db 2046 CACTGTACACTCGCTTTCAGCAGCTGTGGAGAGCTGTTACAAAGGTTGAGAGCA 2104
QY 2106 TTGGGATGCTTGTGAGTGGGGATTTACAAAGAGAGTGTACTCTCAGCTGCTTC 2165
Db 2106 TTGGGATGCTTGTGAGTGGGGATTTACAAAGAGAGTGTACTCTCAGCTGCTTC 2164
QY 2166 AGATAGCGTGTGTCGCTTGAAGATGATGAGCTCTCTGATGAAGTATGTCCTCCCATG 2225
Db 2166 AGATAGCGTGTGTCGCTTGAAGATGATGAGCTCTCTGATGAAGTATGTCCTCCCATG 2224

Qy	426	AGTGTCTCTTGAATTGATAGTGGCTTCCTGTTTGTCAAGTCTCATATAAGAACTTACAGCTCAT	485
Db	425	AGTGTCTCTTGAATTGATAGTGGCTTCCTGTTTGTCAAGTCTCATATAAGAACTTACAGCTCAT	484
Qy	486	CAGGAGAGATCGCAGCAGGGTAGAGACACCAACACACATGTTCTGCACGAGCTCAAGG	545
Db	485	CAGGAGAGATCGCAGCAGGGTAGAGACACCAACACATGTTCTGCACGAGCTCAAGG	544
Qy	546	ATCTCAAGATCACAGGAGTGTCTTCTCTCTTACTGGCACCAAGGTCAGTTCCTCAACG	605
Db	545	ATCTCAAGATCACAGGAGTGTCTTCTCTCTTACTGGCACCAAGGTCAGTTCCTCAACG	604
Qy	606	AGTCTTCAGAGGAGCAGCAGGAAGCTCAGAGAGCTGCAAAAGCAACCGTGCCCATCTGTC	665
Db	605	AGTCTTCAGAGGAGCAGCAGGAAGCTCAGAGAGCTGCAAAAGCAACCGTGCCCATCTGTC	664
Qy	666	AAGACATCTCTGAGAAGAACATACAAAGAAAGTCTTCTCTCAAGGAAAACCAAGTCGGAGCC	725
Db	665	AAGACATCTCTGAGAAGAACATACAAAGAAAGTCTTCTCTCAAGGAAAACCAAGTCGGAGCC	724
Qy	726	GAGTCTATCTTCACACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCCAGAGTTTGAAC	785
Db	725	GAGTCTATCTTCACACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCCAGAGTTTGAAC	784
Qy	786	GGCTGAATGTTGCATCTCAGAGAACATTTGGCAAGACACAAATATAAAGAAAGCAGGAAT	845
Db	785	GGCTGAATGTTGCATCTCAGAGAACATTTGGCAAGACACAAATATAAAGAAAGCAGGAAT	844
Qy	846	CTTTGGAAAAGAGAGACTTTTGA AAAACAATTTGCAGAGCAAGCAAGTTTGCAGCAGAGTTC	905
Db	845	CTTTGGAAAAGAGAGACTTTTGA AAAACAATTTGCAGAGCAAGCAAGTTTGCAGCAGAGTTC	904
Qy	906	CAGTGGAGGTTATCAAAAGATCTCTTGGTGAAGAGGTTTTTAAATATGTTACAGGAAG	965
Db	905	CAGTGGAGGTTATCAAAAGATCTCTTGGTGAAGAGGTTTTTAAATATGTTACAGGAAG	964
Qy	966	ATGAAACATCTTTGGGGTGGTGGAGGCACCTTTAAAGATTTTTTAAACAGCTTCAGTA	1025
Db	965	ATGAAACATCTTTGGGGTGGTGGAGGCACCTTTAAAGATTTTTTAAACAGCTTCAGTA	1024
Qy	1026	CCCTTCTGAAACAGAGCAGCCATTTGCCAAGAGCAGGAAAAGGGCAGGCTTCAGGACG	1085
Db	1025	CCCTTCTGAAACAGAGCAGCCATTTGCCAAGAGCAGGAAAAGGGCAGGCTTCAGGACG	1084
Qy	1086	CCCTCATCTATGCCCTGGATAAGAGAGTATTTTCTACATGTTTACTTACTTCTTCCTTA	1145
Db	1085	CCCTCATCTATGCCCTGGATAAGAGAGTATTTTCTACATGTTTACTTACTTCTTCCTTA	1144
Qy	1146	AGAGAACCCCTCCCTGATTTCTCCCGGCATCATAAAGGCAGCTGCTCAGTATTATATG	1205
Db	1145	AGAGAACCCCTCCCTGATTTCTCCCGGCATCATAAAGGCAGCTGCTCAGTATTATATG	1204
Qy	1206	AAACGGAAAGTGGAGTTCGCTTAATGCTCCCTGCTTCCATTAATGATTGACGAGTTTG	1265
Db	1205	AAACGGAAAGTGGAGTTCGCTTAATGCTCCCTGCTTCCATTAATGATTGACGAGTTTG	1264
Qy	1266	TGAATCAGCCCTACTTGTGTACTCCCTTGCATGAAAGACACCAAGCCATCCCTGTCCC	1325
Db	1265	TGAATCAGCCCTACTTGTGTACTCCCTTGCATGAAAGACACCAAGCCATCCCTGTCCC	1324
Qy	1326	CCAGCAACCCCTCCTCGCTGGTGATTTCCACATCGCTATTCTGCAAGACATTTCCAT	1385
Db	1325	CCAGCAACCCCTCCTCGCTGGTGATTTCCACATCGCTATTCTGCAAGACATTTCCAT	1384
Qy	1386	TCCATTTTCATGTTGACAAAAGATGACAAATCTTGCAATTTGGCAATGGCATCAGAAGGC	1445
Db	1385	TCCATTTTCATGTTGACAAAAGATGACAAATCTTGCAATTTGGCAATGGCATCAGAAGGC	1444
Qy	1446	TGATGAACAGGAGACTTTCAAGGAAGCCATAATTTTGAAGAAATACTTTGAAATCTCGA	1505
Db	1445	TGATGAACAGGAGACTTTCAAGGAAGCCATAATTTTGAAGAAATACTTTGAAATCTCGA	1504

QY	1506	CTCCAAAATCAACCGACAGCGTTTACGGGGATCATGACTATGTTGAATATATGCAAGTTTCTTG	1561
DB	1505	CTCCAAAATCAACCGACAGCGTTTACGGGGATCATGACTATGTTGAATATATGCAAGTTTCTTG	1564
QY	1566	TACGAGTGAGGAGATGGACAACCTCTGTGAAGAAATCTTCAAGGGTTTATGGACCTCAAG	1625
DB	1565	TACGAGTGAGGAGATGGACACACTCTGTGAAGAAATCTTCAAGGGTTTATGGACCTCAAG	1624
QY	1626	GCCAAATGANTACATTGTTGAATCCAGTGCAAATCTTGTGTTTGGGGTACCCTGTGTGG	1685
DB	1625	GCCAAATGANTACATTGTTGAATCCAGTGCAAATCTTGTGTTTGGGGTACCCTGTGTGG	1684
QY	1686	ACAGATTAGAGATTTTACAGGAGAGGGCTTCACTCTCAGACATCCCAATTCACATG	1745
DB	1685	ACAGATTAGAGATTTTACAGGAGAGGGCTTCACTCTCAGACATCCCAATTCACATG	1744
QY	1746	CACGTAGGGATGTGGTCTTAATAGGGGAACAAGCCGAGCTCAAGATGGCGCTCAAGAAGA	1805
DB	1745	CACGTAGGGATGTGGTCTTAATAGGGGAACAAGCCGAGCTCAAGATGGCGCTCAAGAAGA	1804
QY	1806	GGCTGGGAAGCTGAAGGCTTACCTTGAGCAAGCCCAACCAAGCCCTGGAGGAGGAGAAGA	1865
DB	1805	GGCTGGGAAGCTGAAGGCTTACCTTGAGCAAGCCCAACCAAGCCCTGGAGGAGGAGAAGA	1864
QY	1866	AAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTGCTCAGCAGCTGTGGC	1925
DB	1865	AAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTGCTCAGCAGCTGTGGC	1924
QY	1926	AAGGGCAAGTTGTCAAGCAAGCAAGTTCACTAATGTCACCATGCTCTTCTCAGACATCG	1985
DB	1925	AAGGGCAAGTTGTCAAGCAAGCAAGTTCACTAATGTCACCATGCTCTTCTCAGACATCG	1984
QY	1986	TTGGGTTTCACTGCCATCTGCTCCAGTGCACCGCTGCAGGTATCATCACCATGCTCAATG	2045
DB	1985	TTGGGTTTCACTGCCATCTGCTCCAGTGCACCGCTGCAGGTATCATCACCATGCTCAATG	2044
QY	2046	CACGTATACCTCGTTTCCAGCAGCAGTGTGGAGAGCTGATGTCATCAAGGTTGGAGACCA	2105
DB	2045	CACGTATACCTCGTTTCCAGCAGCAGTGTGGAGAGCTGATGTCATCAAGGTTGGAGACCA	2104
QY	2106	TTGGGATGGCTATTTGTGTAAGCTGGGGATTCACAAAGAGAGTGATCATGCTGTTTC	2165
DB	2105	TTGGGATGGCTATTTGTGTAAGCTGGGGATTCACAAAGAGAGTGATCATGCTGTTTC	2164
QY	2166	AGATAGGGCTGATGGCCCTCAAGATGATGAGGTCTCTGATGAAGTTATGTCCTCCCATG	2225
DB	2165	AGATAGGGCTGATGGCCCTCAAGATGATGAGGTCTCTGATGAAGTTATGTCCTCCCATG	2224
QY	2226	GAGAACCTATCAAGATCGGAATTTGGACTGCACCTTGATCAGTTTTTGTGGCGTGTGG	2285
DB	2225	GAGAACCTATCAAGATCGGAATTTGGACTGCACCTTGATCAGTTTTTGTGGCGTGTGG	2284
QY	2286	GAGTTAAATGCCCGCTTACTGTCTTTTGGAAACAATGTCCTGTGGCTTAACAAATTTG	2345
DB	2285	GAGTTAAATGCCCGCTTACTGTCTTTTGGAAACAATGTCCTGTGGCTTAACAAATTTG	2344
QY	2346	AGTCTCGAGTGATACAGAAAATCAATGTACGCCCAACAACCTTACAGATTTACTCAAAG	2405
DB	2345	AGTCTCGAGTGATACAGAAAATCAATGTACGCCCAACAACCTTACAGATTTACTCAAAG	2404
QY	2406	ACTGTCTGGTTTGGTGTGTTTACCCCTCGATCAAGGGAGGAATTCACACCAACTTCCCTA	2465
DB	2405	ACTGTCTGGTTTGGTGTGTTTACCCCTCGATCAAGGGAGGAATTCACACCAACTTCCCTA	2464
QY	2466	GTGAATCCCGGAATCTGCCATTTTCTTGGATGCTTACCACAAAGGAACAACTCAAAAC	2525
DB	2465	GTGAATCCCGGAATCTGCCATTTTCTTGGATGCTTACCACAAAGGAACAACTCAAAAC	2524
QY	2526	CATGCTTCCAAAAGAAAGTGTGGAAGATGGCAATGCCAATTTTTTATAGGCAAGCATCAG	2585
DB	2525	CATGCTTCCAAAAGAAAGTGTGGAAGATGGCAATGCCAATTTTTTATAGGCAAGCATCAG	2584
QY	2586	GAATAGATTTAGCAACCTATATACCTATTTTAAAGTCTTTTGGGTTTGTACTCATTTGAAGAT	2645

Db 2585 GAATAGATTAGCAACCTATATACCTATTATTAAGTCTTTGGGTTTGACTCATTTGAAGAT 2644
 QY 2646 GTGTAGAGCCTCTGAAGCACTTTAGGGATTGTAGATGCTTAACAGCAGTAGTATTAAT 2705
 Db 2645 GTGTAGAGCCTCTGAAGCACTTTAGGGATTGTAGATGCTTAACAGCAGTAGTATTAAT 2704
 QY 2706 TCAGGAGCAAGTCACAACTTTCTCTCTTTAAACATGACAAATGTACTACTCAGTA 2765
 Db 2705 TCAGGAGCAAGTCACAACTTTCTCTCTTTAAACATGACAAATGTACTACTCAGTA 2764
 QY 2766 CTTACAGCTCTCAAGAAAAAACCCTTAAAGAGCTACTTTTGTGGAGTATTCT 2825
 Db 2765 CTTACAGCTCTCAAGAAAAAACCCTTAAAGAGCTACTTTTGTGGAGTATTCT 2824
 QY 2826 ATTATATACCAAGCACTTACTACCTGTACTCAAAATTCAGCACTTTGTATATATCAGA 2885
 Db 2825 ATTATATACCAAGCACTTACTACCTGTACTCAAAATTCAGCACTTTGTATATATCAGA 2884
 QY 2886 TAATTGTAGTCAATTGTACAACTGATGAGTCACCTGCAATCTCATATCTCTGGTGAAT 2945
 Db 2885 TAATTGTAGTCAATTGTACAACTGATGAGTCACCTGCAATCTCATATCTCTGGTGAAT 2944
 QY 2946 GCATGGTTATTAAAGTGTGTTGTGATAGTGTCTCAAAAAAACCCTTAAAGAGTATTCT 3005
 Db 2945 GCATGGTTATTAAAGTGTGTTGTGATAGTGTCTCAAAAAAACCCTTAAAGAGTATTCT 3004
 QY 3006 AAAAAA 3011
 Db 3005 ATTACA 3010

RESULT 4

ABV27167
 ID ABV27167 standard; cDNA; 4742 BP.

XX AC ABV27167;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cdna 27158.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US05171.

XX PR 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 5509-5510; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (i) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX

SQ Sequence 4742 BP; 1450 A; 933 C; 998 G; 1355 T; 6 other;

Query Match 97.9%; Score 2950.8; DB 23; Length 4742;
 Best Local Similarity 99.1%; Pred. No. 0;
 Matches 2978; Conservative 0; Mismatches 27; Indels 1; Gaps 1;
 QY 6 ATGGCGATTGGCGGCTGCAGAGACCAGGACTCAGTTCCCTGCCCTAGTCTCAGGCTAG 65
 Db 6 ATGGCGATTGGCGGCTGCAGAGACCAGGACTCAGTTCCCTGCCCTAGTCTCAGGCTAG 65
 QY 66 TGGTGGGACTCAGCTCAGAGTCAGTTTCCAGAGCAGAGTTTCAGTGCAGAGTTTTCCT 125
 Db 66 TGGTGGGACTCAGCTCAGAGTCAGTTTCCAGAGCAGAGTTTCAGTGCAGAGTTTTCCT 125
 QY 126 ACACCTTTTCTCGCTAGAGCAGCAGCAGCTGGAACAGCCAGCCAGCGGAGGACACCTG 185
 Db 126 ACACCTTTTCTCGCTAGAGCAGCAGCAGCTGGAACAGCCAGCGGAGGACACCTG 185
 QY 186 TGGGGAGGAGCGCCTGGAGGAGCTTAGAGACCCAGCGGCGGTGATCTACCATGTG 245
 Db 186 TGGGGAGGAGCGCCTGGAGGAGCTTAGAGACCCAGCGGCGGTGATCTACCATGTG 245
 QY 246 CGGATTTCAGAGCGCGCCTGGAGCTGTAGAGATCCGGAAGCAGACGCCAGGTTG 305
 Db 246 CGGATTTCAGAGCGCGCCTGGAGCTGTAGAGATCCGGAAGCAGACGCCAGGTTG 305
 QY 306 CGAAGCCACCAAGACTGCGGCTCTTGGAGAAACCGTGACAGGGGCGCCACCGGTCCTC 365
 Db 306 CGAAGCCACCAAGACTGCGGCTCTTGGAGAAACCGTGACAGGGGCGCCACCGGTCCTC 365
 QY 366 GCGCCTGTCTGCACCCCTGTGCGCTGAGCTGCCTGACAGTGACATGACATCCAGTTACC 425
 Db 366 GCGCCTGTCTGCACCCCTGTGCGCTGAGCTGCCTGACAGTGACATGACATCCAGTTACC 424
 QY 426 AGTGTCTTGAATTGATAGTGGCTTCTGTTGTGCTCTCATATAGAACTACAGTCTAT 485
 Db 426 AGTGTCTTGAATTGATAGTGGCTTCTGTTGTGCTCTCATATAGAACTACAGTCTAT 484
 QY 486 CAGGAGGAGATCGCAGCAGGGTAAGAGACACCAACCACTGTTCTCCAGCAAGCTCAAGG 545
 Db 486 CAGGAGGAGATCGCAGCAGGGTAAGAGACACCAACCACTGTTCTCCAGCAAGCTCAAGG 544
 QY 546 ATCTCAAGATCACAGGAGAGTGTCTTCTCTTACTGGCACCAGGTCAAGTTCTCAACG 605
 Db 546 ATCTCAAGATCACAGGAGAGTGTCTTCTCTTACTGGCACCAGGTCAAGTTCTCAACG 604
 QY 606 AGTCTTCAGAGGAGGAGGAGGAGCTCAGAGAGTGCAGAGCAACCGTGCCTCATCTGTC 665
 Db 606 AGTCTTCAGAGGAGGAGGAGGAGCTCAGAGAGTGCAGAGCAACCGTGCCTCATCTGTC 664
 QY 666 AAGACATTCTCAGAGAGAACATACAAAGTCTTCTCTCAAGAAAAAACCAGTCSGAGCC 725
 Db 666 AAGACATTCTCAGAGAGAACATACAAAGTCTTCTCTCAAGAAAAAACCAGTCSGAGCC 724
 QY 726 GAGTCTATCTTACACTTTTGGCAGAGAGTATTTCGCAACTGATTTTCCAGAGTTTGAAC 785
 Db 726 GAGTCTATCTTACACTTTTGGCAGAGAGTATTTCGCAACTGATTTTCCAGAGTTTGAAC 784

QY 786 GGCTGAATGTTGCATCTTCAGAGAACATATGCGCAAGACACAAATAAAGAAAGCAGGAAAT 845
DB 785 GGCTGAATGTTGCATCTTCAGAGAACATATGCGCAAGACACAAATAAAGAAAGCAGGAAAT 844
QY 846 CTTTGGAAAGAGAGACTTTTGAAGAAACATTCAGAGCAAGCAGTTCAGCAGGAGTTC 905
DB 845 CTTTGGAAAGAGAGACTTTTGAAGAAACATTCAGAGCAAGCAGTTCAGCAGGAGTTC 904
QY 906 CAGTGGAGGTTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAATATGTTTACGAGGAAG 965
DB 905 CAGTGGAGGTTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAATATGTTTACGAGGAAG 964
QY 966 ATGAAACATCTCTTGGGGTGGTGGAGGCCCTTAAAGATTTTTTAAACAGCTTCAGTA 1025
DB 965 ATGAAACATCTCTTGGGGTGGTGGAGGCCCTTAAAGATTTTTTAAACAGCTTCAGTA 1024
QY 1026 CCCTCTGAAACAGAGCAGCATTCGCAAGAACAGGAAAGGGCAGGCTTGAGGACG 1085
DB 1025 CCCTCTGAAACAGAGCAGCATTCGCAAGAACAGGAAAGGGCAGGCTTGAGGACG 1084
QY 1086 CCTCCATTTCTATGCTGGTAAGAGGATGATTTTCTACATGTTTACTACTTCTTCCCTA 1145
DB 1085 CCTCCATTTCTATGCTGGTAAGAGGATGATTTTCTACATGTTTACTACTTCTTCCCTA 1144
QY 1146 AGAAGACCCCTCCCTGATTTCTCCGGGCATCATAAAGCAGCTGCTCAGTATATATG 1205
DB 1145 AGAAGACCCCTCCCTGATTTCTCCGGGCATCATAAAGCAGCTGCTCAGTATATATG 1204
QY 1206 AAACGGAAGTGGAGTCTGCTTAATGCTCCCTGCTTCCATAATGATGTCAGCGAGTTG 1265
DB 1205 AAACGGAAGTGGAGTCTGCTTAATGCTCCCTGCTTCCATAATGATGTCAGCGAGTTG 1264
QY 1266 TGAATCAGCCCTACTTGTGTTACTCCGTTTACATGAAAGCAGCAGCCATCCCTGTCCTC 1325
DB 1265 TGAATCAGCCCTACTTGTGTTACTCCGTTTACATGAAAGCAGCAGCCATCCCTGTCCTC 1324
QY 1326 CCAGCAAAACCCAGTCTCCCTGGTATGCCACATGCTTATCTGCAAGACATATTCAT 1385
DB 1325 CCAGCAAAACCCAGTCTCCCTGGTATGCCACATGCTTATCTGCAAGACATATTCAT 1384
QY 1386 TCCATTTCTATGTTGACAAAGATGATGCAATTTGCAATTTGGCAATGGCATCAGAGGC 1445
DB 1385 TCCATTTCTATGTTGACAAAGATGATGCAATTTGCAATTTGGCAATGGCATCAGAGGC 1444
QY 1446 TGATGAACAGGAGAGACTTTCAAGGAAGCCTTAATTTGAAGAAATACTTTGAAATCTGA 1505
DB 1445 TGATGAACAGGAGAGACTTTCAAGGAAGCCTTAATTTGAAGAAATACTTTGAAATCTGA 1504
QY 1506 CTCCAAAATCAACACAGACGTTTACGGGATCATGACTATGTTGAATATGCACTTTGTTG 1565
DB 1505 CTCCAAAATCAACACAGACGTTTACGGGATCATGACTATGTTGAATATGCACTTTGTTG 1564
QY 1566 TACAGTGAAGAGATGGGCAACTCTGTGAAGAAATCTTCAAGGGTTATGGACCTCAAG 1625
DB 1565 TACAGTGAAGAGATGGGCAACTCTGTGAAGAAATCTTCAAGGGTTATGGACCTCAAG 1624
QY 1626 GCCAAATGATCTACATTTGTAATCCAGTGCATCTTGTGTTTTGGGCTACCCCTGTGTGG 1685
DB 1625 GCCAAATGATCTACATTTGTAATCCAGTGCATCTTGTGTTTTGGGCTACCCCTGTGTGG 1684
QY 1686 ACAGATTAGAAGATTTTACAGGAGGAGGCTTACCTCTCAGACATCCCAATTCACATG 1745
DB 1685 ACAGATTAGAAGATTTTACAGGAGGAGGCTTACCTCTCAGACATCCCAATTCACATG 1744
QY 1746 CACTGAGGAGTGGTCTTAATAGGGGAACAAGCCGAGCTCAAGATGGCTCAAGAGA 1805
DB 1745 CACTGAGGAGTGGTCTTAATAGGGGAACAAGCCGAGCTCAAGATGGCTCAAGAGA 1804
QY 1806 GGCTGGGAAGCTGAAGGCTTACCTTGAAGCAAGCCACCAAGCCCTGGAGGAGAGA 1865
DB 1805 GGCTGGGAAGCTGAAGGCTTACCTTGAAGCAAGCCACCAAGCCCTGGAGGAGAGA 1864
QY 1866 AAAAGACAGTAGACCTTCTGCTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCTGTGGC 1925

DB 1865 AAAAGACAGTAGACCTTCTGCTGCTCCATATTTCCCTGTAGGTTGCTCAGCAGCTGTGGC 1924
QY 1926 AAGGCAAGTTGTGCAAGCCCAAGTTCAGTATGTCCACCATGCTCTCTCAGACATCG 1985
DB 1925 AAGGCAAGTTGTGCAAGCCCAAGTTCAGTATGTCCACCATGCTCTCTCAGACATCG 1984
QY 1986 TTTGGGTTCACTGCTGCTCCAGTGTCCAGGTGCTCAGGTCATCACCATTGCTCAATG 2045
DB 1985 TTTGGGTTCACTGCTGCTCCAGTGTCCAGGTGCTCAGGTCATCACCATTGCTCAATG 2044
QY 2046 CACTGTACACTGCTCTCGACACAGTGTGGAGAGCTGGAGTGTCTTCAAGGTGGAGACCA 2105
DB 2045 CACTGTACACTGCTCTCGACACAGTGTGGAGAGCTGGAGTGTCTTCAAGGTGGAGACCA 2104
QY 2106 TTTGGCATGCTTATGTTAGTGTGGGATTCACAAAGAGAGTGTACTCATGCTGTTC 2165
DB 2105 TTTGGCATGCTTATGTTAGTGTGGGATTCACAAAGAGAGTGTACTCATGCTGTTC 2164
QY 2166 AGATAGCGCTGATGGCCCTGAAGATGATGGAGCTCTCTGATGAAGTATGCTCCTCCCATG 2225
DB 2165 AGATAGCGCTGATGGCCCTGAAGATGATGGAGCTCTCTGATGAAGTATGCTCCTCCCATG 2224
QY 2226 GAGAACTATCAAGATGCGAATTTGGAGTGTGCTGATCAGTTTTTCTGGGCTGTGTTG 2285
DB 2225 GAGAACTATCAAGATGCGAATTTGGAGTGTGCTGATCAGTTTTTCTGGGCTGTGTTG 2284
QY 2286 GAGTTAAATGCCCCGTTACTGTCTTTTGGAAACAATGTCCTCTGGCTTAACAATAATG 2345
DB 2285 GAGTTAAATGCCCCGTTACTGTCTTTTGGAAACAATGTCCTCTGGCTTAACAATAATG 2344
QY 2346 AGTCTGCAAGTGTACACGAAATAATCAATGTACGCCCAACAACCTTACAGATTACTCAAAG 2405
DB 2345 AGTCTGCAAGTGTACACGAAATAATCAATGTACGCCCAACAACCTTACAGATTACTCAAAG 2404
QY 2406 ACTGTCTGTTGTTGTTTACCCCTCGATCAAGGAGGAGAACTTCCACCAAACTTCCTTA 2465
DB 2405 ACTGTCTGTTGTTGTTTACCCCTCGATCAAGGAGGAGAACTTCCACCAAACTTCCTTA 2464
QY 2466 GTCAAAATCCCGGAATCTGCAATTTTCTGGATGCTTACCAACAAGGAACAACCTCAAAAC 2525
DB 2465 GTCAAAATCCCGGAATCTGCAATTTTCTGGATGCTTACCAACAAGGAACAACCTCAAAAC 2524
QY 2526 CATGCTTCCAAAAGAAGATGTGGAAGATGGCAATGCCAATTTTTAGGCAAGCATCAG 2585
DB 2525 CATGCTTCCAAAAGAAGATGTGGAAGATGGCAATGCCAATTTTTAGGCAAGCATCAG 2584
QY 2586 GAATAGATTAGCAACCTATATACCTATTTATAGTCTTTGGGGTTTGACTCATTTGAAGAT 2645
DB 2585 GAATAGATTAGCAACCTATATACCTATTTATAGTCTTTGGGGTTTGACTCATTTGAAGAT 2644
QY 2646 GTGTAGAGCTCTGAAAGCACTTTAGGGATTGTAGATGGCTAAACAAGCAGTATTAATAAT 2705
DB 2645 GTGTAGAGCTCTGAAAGCACTTTAGGGATTGTAGATGGCTAAACAAGCAGTATTAATAAT 2704
QY 2706 TCAGGAGCCCAAGTCACAATCTTTCTCCTGTTTAAACATGACAAAATGTACTCACTTCAGTA 2765
DB 2705 TCAGGAGCCCAAGTCACAATCTTTCTCCTGTTTAAACATGACAAAATGTACTCACTTCAGTA 2764
QY 2766 CTTTCAGCTCTTCAAGAAAAAACCCTTAAAAAGCTACTTTTGTGGGAGTATTTCT 2825
DB 2765 CTTTCAGCTCTTCAAGAAAAAACCCTTAAAAAGCTACTTTTGTGGGAGTATTTCT 2824
QY 2826 ATTATATAACAGCAGCTTACTACTGTACTCAAAATTCAGCAGCTTGTACATATATCAGA 2885
DB 2825 ATTATATAACAGCAGCTTACTACTGTACTCAAAATTCAGCAGCTTGTACATATATCAGA 2884
QY 2886 TAATTGTAGTCAATTTTACAACTGATGGAGTCACCTGCAATCTCATATCTCTGTTGGAAT 2945
DB 2885 TAATTGTAGTCAATTTTACAACTGATGGAGTCACCTGCAATCTCATATCTCTGTTGGAAT 2944
QY 2946 GCCATGGTTATTAAGTGTGTTTGTGATAGTGTGTCACAAAAAATAAAAAAATAAAAAA 3005

[illegible]

Db	2225	GAGAACCTATCAAGATCGGAATTTGGAGCTGCACCTCTGGATCAGTTTTTGTCTGGCGTCTGTTG	2288
Qy	2286	GAGTTAAATGCCCGCTTACTGTCTTTTGGAAACAATGTCACCTCTGGCTTAACAAATTTG	2345
Db	2285	GAGTTAAATGCCCGCTTACTGTCTTTTGGAAACAATGTCACCTCTGGCTTAACAAATTTG	2344
Qy	2346	AGTCCTCCAGTGTACCAACGAAAAATCAATGTCAGCCCCAACAACTTACAGATTACTCAAA	2405
Db	2345	AGTCCTCCAGTGTACCAACGAAAAATCAATGTCAGCCCCAACAACTTACAGATTACTCAAA	2404
Qy	2406	ACTGTCTCGTTTCGTTTACCCCTCGATCAGGGAGGAACCTCCACCAAACTTCCTTA	2465
Db	2405	ACTGTCTCGTTTCGTTTACCCCTCGATCAGGGAGGAACCTCCACCAAACTTCCTTA	2464
Qy	2466	GTGAAATCCCGGAATCTGCAATTTTCTGGATGCTTACCAACAAGGAACAACTCAAAAC	2525
Db	2465	GTGAAATCCCGGAATCTGCAATTTTCTGGATGCTTACCAACAAGGAACAACTCAAAAC	2524
Qy	2526	CATGCTTCCAAAAAGAAAGATGTGGAAGATGGCAATGCCAATTTTTTAGCCAAAGCATCAG	2585
Db	2525	CATGCTTCCAAAAAGAAAGATGTGGAAGATGGCAATGCCAATTTTTTAGCCAAAGCATCAG	2584
Qy	2586	GAATAGATTAGCAACCTATATACCTATTATATAGTCTTTGGGGTTTGACTCATTTGAAGAT	2645
Db	2585	GAATAGATTAGCAACCTATATACCTATTATATAGTCTTTGGGGTTTGACTCATTTGAAGAT	2644
Qy	2646	GTGTAGAGCCTCTGAAAGCACTTTAGGGATTGTAGATGGCTTAACAAGCAGTATTAATAAT	2705
Db	2645	GTGTAGAGCCTCTGAAAGCACTTTAGGGATTGTAGATGGCTTAACAAGCAGTATTAATAAT	2704
Qy	2706	TCAGGAGCCAAAGTCACAACTTTCTCTGCTTTAACATGACAAAATGTACTCACTTCAGTA	2765
Db	2705	TCAGGAGCCAAAGTCACAACTTTCTCTGCTTTAACATGACAAAATGTACTCACTTCAGTA	2764
Qy	2766	CTTCAGCTCTTCAAGAAAAAATAAAACCTTAAAAGCTACTTTTGTGGAGTATTTCT	2825
Db	2765	CTTCAGCTCTTCAAGAAAAAATAAAACCTTAAAAGCTACTTTTGTGGAGTATTTCT	2824
Qy	2826	ATTATATAACGACACTTACTACTGTACTCAAAATTCAGCACCTTGACATATATCAGA	2885
Db	2825	ATTATATAACGACACTTACTACTGTACTCAAAATTCAGCACCTTGACATATATCAGA	2884
Qy	2886	TAATTGTAGTCAATTGTACAACTGATGGAGTCACCTGCAATCTCATATCTCTGGTGAAT	2945
Db	2885	TAATTGTAGTCAATTGTACAACTGATGGAGTCACCTGCAATCTCATATCTCTGGTGAAT	2944
Qy	2946	GCCATGTTTATTAAGTGTGTTTGTGATAGTGTGTCGCAAAAAAATAAAAAA	3005
Db	2945	GCCATGTTTATTAAGTGTGTTTGTGATAGTGTGTCGCAAAAAAATAAAAAA	3004
Qy	3006	AAAAA 3011	
Db	3005	ATTAACA 3010	
RESULT 6			
ABK92193			
ID	ABK92193 standard; DNA; 3004 BP.		
XX			
AC	ABK92193;		
XX			
DT	15-AUG-2002 (first entry)		
XX			
DE	Prostate cancer-associated DNA sequence #79.		
XX			
KW	Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;		
KW	gene therapy; gene; ds.		
XX			
OS	Mammalia.		
XX			
PN	WO200230268-A2.		
XX			
PD	18-APR-2002.		

Db 1374 TCCATTCATTTTCATGTTTCACAAAGATATGACAAATTCGCAATTTGGCAATGGCATCAG 1433
QY 1441 AAGSCTGATGAACAGAGAGACTTTCAGGAAGCCCTAAATTTTGAAGAACTACTTTGAAAT 1500
Db 1434 AAGSCTGATGAACAGAGAGACTTTCAGGAAGCCCTAAATTTT --GAATACITTTGAAT 1490
QY 1501 TCTGACTCCAAAATCAACCAGAGGTTTACGGGATCATGACTATGTTGAATATGCAATT 1560
Db 1491 TCTGACTCCAAAATCAACCAGAGACTTTCAGGGGATCATGACTATGTTGAATATGCAATT 1550
QY 1561 TGTGTACGAGTGAAGAGATGGGACAACTCTGTGAAGAAATCTTCAAGGGTATGGACCT 1620
Db 1551 TGTGTACGAGTGAAGAGATGGGACAACTCTGTGAAGAAATCTTCAAGGGTATGGACCT 1610
QY 1621 CAAGGCCAAATGATCTACATCTTGTGAATCCAGTGCATCTGTGTTTGGGGTCAACCTG 1680
Db 1611 CAAGGCCAAATGATCTACATCTTGTGAATCCAGTGCATCTGTGTTTGGGGTCAACCTG 1670
QY 1681 TGTGGACAGATTAAGAGATTTTACAGCAGAGGCTCTACCTCTCGACATATCCCAATTC 1740
Db 1671 TGTGGACAGATTAAGAGATTTTACAGCAGAGGCTCTACCTCTCGACATATCCCAATTC 1730
QY 1741 CAATGCACTGAGGATGTGCTCTTAATAGGGGAACAGCCGAGCTCAAGATGSCCTGAA 1800
Db 1731 CAATGCACTGAGGATGTGCTCTTAATAGGGGAACAGCCGAGCTCAAGATGSCCTGAA 1790
QY 1801 GAAGAGCTGGGAAGCTGAAGCTACCCCTTGAGCAAGCCCAAGCCCTGGAGGAGGA 1860
Db 1791 GAAGAGCTGGGAAGCTGAAGCTACCCCTTGAGCAAGCCCAAGCCCTGGAGGAGGA 1850
QY 1861 GAAGAAAACAGATAGACCTTCTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCT 1920
Db 1851 GAAGAAAACAGATAGACCTTCTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCT 1910
QY 1921 GTGCAAGGCAAGTGTGCAAGCAAGAAAGTTTCAAGTATGTCACCATGCTCTTCTCAGA 1980
Db 1911 GTGCAAGGCAAGTGTGCAAGCAAGAAAGTTTCAAGTATGTCACCATGCTCTTCTCAGA 1970
QY 1981 CATCGTTGGGTTTCACTGCCATCTGCTCCAGTGTCTCACCCTGACGCTCATCACCATGCT 2040
Db 1971 CATCGTTGGGTTTCACTGCCATCTGCTCCAGTGTCTCACCCTGACGCTCATCACCATGCT 2030
QY 2041 CAATGCACTACACGCTTCGACCAAGCAGTGTGGAGAGCTGGATGCTTACAAAGTGGA 2100
Db 2031 CAATGCACTACACGCTTCGACCAAGCAGTGTGGAGAGCTGGATGCTTACAAAGTGGA 2090
QY 2101 GACCATTTGGCGATGCCCTATTTGCTGAGC - TGGGGGATTACAAAGAGAGTACTCATG 2159
Db 2091 GACCATTTGGCGATTTGCTGCTTGGGGATTACAAAGAGAGTACTCATG 2149
QY 2160 CTGTTGAGATGAGCGCTGATGGCCCTGAAGATGATGAGAGCTCTCTGATGAAGTTATGCTC 2219
Db 2150 CTGTTGAGATGAGCGCTGATGGCCCTGAAGATGATGAGAGCTCTCTGATGAAGTTATGCTC 2209
QY 2220 CCCATGAGAACTATCAAGATGCGAATTTGGACTGCACCTGGATCAGTTTTTGTCTGGCG 2279
Db 2210 CCCATGAGAACTATCAAGATGCGAATTTGGACTGCACCTGGATCAGTTTTTGTCTGGCG 2269
QY 2280 TCGTTGAGTTAAATGCCCGCTTACTGTCTTTTGGAAACAATGTCACCTGCTGGCTAAACA 2339
Db 2270 TCGTTGAGTTAAATGCCCGCTTACTGTCTTTTGGAAACAATGTCACCTGCTGGCTAAACA 2329
QY 2340 AATTTGAGTCTGAGTGTACCAAGAAAATCAATGTACGCCCAACAACATACAGATTAC 2399
Db 2330 AATTTGAGTCTGAGTGTACCAAGAAAATCAATGTACGCCCAACAACATACAGATTAC 2389
QY 2400 TCAAGACTCTCTGTTGCTTTTACCCCTCGATCAAGGGAGGAACTTCCACCAAACT 2459
Db 2390 TCAAGACTCTCTGTTGCTTTTACCCCTCGATCAAGGGAGGAACTTCCACCAAACT 2449
QY 2460 TCCTAGTAAATCCCGGAATCTGCAATTTTCTGGATGCTTACCAAGAGAACT 2519
Db 2450 TCCTAGTAAATCCCGGAATCTGCAATTTTCTGGATGCTTACCAAGAGAACT 2509

QY 2520 CAAAACCATGCTTCCAAAAGAAAAGATGTGGAAGATGCAATTTTGGCAAAAG 2579
Db 2510 CAAAACCATGCTTCCAAAAGAAAAGATGTGGAAGATG - CAAGCAATTTTGGCAAAAG 2567
QY 2580 CATCAGAAATAGATTTAGCAACCTATATACCTATTTATAAGCTTTTGGGGTTTGAATCAT 2639
Db 2568 CATCAGAAATAGATTTAGCAACCTATATACCTATTTATAAGCTTTTGGGGTTTGAATCAT 2627
QY 2640 GAAGATGTGAGAGCCTCTGAAAAGCACTTTAGGATTTGTAGTGGCTTAAACAGCATTT 2699
Db 2628 GAAGATGTGAGAGCCTCTGAAAAGCACTTTAGGATTTGTAGTGGCTTAAACAGCATTT 2687
QY 2700 AAAATTTTCAGAGCCCAAGTCAACATCTTTCTCTGTTTAAACATGACAAAATGTACTCACT 2759
Db 2688 AAAATTTTCAGAGCCCAAGTCAACATCTTTCTCTGTTTAAACATGACAAAATGTACTCACT 2747
QY 2760 TCAGTACTTTCAGCTCTTCAAGAAAACCTTAAAGAGCTACTTTTGTGGGAGT 2819
Db 2748 TCAGTACTTTCAGCTCTTCAAGAAAACCTTAAAGAGCTACTTTTGTGGGAGT 2807
QY 2820 ATTTCTATATATAACAGCAGCTTACTACCTGTACTCAAAATTCAGCAGCTTGTACATAT 2879
Db 2808 ATTTCTATATATAACAGCAGCTTACTACCTGTACTCAAAATTCAGCAGCTTGTACATAT 2867
QY 2880 ATCAGATAATTTGATGCAATTTGTACAACTGTAGTGGAGTCACTGCAATCTCATATCTCTGG 2939
Db 2868 ATCAGATAATTTGATGCAATTTGTACAACTGTAGTGGAGTCACTGCAATCTCATATCTCTGG 2927
QY 2940 TGAATGCCATGTTTATTAAGTGTGTTGTGATAG - TGTGCTCAAAAAA 2998
Db 2928 TGAATGCCATGTTTATTAAGTGTGTTGTGATAGTGTGTTGCTCAAAAAA 2987
QY 2999 AAAAAA 3015
Db 2988 AAAAAA 3004

RESULT 7
ABK92259

ID ABK92259 standard; DNA; 3004 BP.

XX AC ABK92259;

XX XX DT

XX 15-AUG-2002 (first entry)

DE Prostate cancer-associated DNA sequence #145.

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
KW gene therapy; gene; ds.

OS Mammalia.

XX PN WO200230268-A2.

XX PD 18-APR-2002.

XX PF 12-OCT-2001; 2001WO-US32045.

XX XX 13-OCT-2000; 2000US-0687576.

PR 08-DEC-2000; 2000US-0733288.

PR 24-JAN-2001; 2001US-263957P.

PR 16-MAR-2001; 2001US-276791P.

PR 06-APR-2001; 2001US-281922P.

PR 24-APR-2001; 2001US-286214P.

PR 30-APR-2001; 2001US-0847046.

PR 04-MAY-2001; 2001US-288589P.

XX XX (BOSB-) EOS BIOTECHNOLOGY INC.

XX PA Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

PI

```
XX WPI: 2002-471335/50.
DR P-PSDB; ABG61940.
XX
PT Detecting a prostate cancer-associated transcript in a cell in a
PT patient, useful for diagnosing prostate cancer (PC) or screening
PT modulators of PC, by determining if prostate cancer-associated genes
PT are expressed in a prostate tissue
XX
PS Claim 22; Page 421; 436pp; English.
XX
XX The present invention relates to methods of detecting a prostate
CC cancer-associated transcript in a cell from a patient. The method
CC comprises contacting a biological sample from the patient with
CC prostate cancer-associated polynucleotides (designated PC genes) that
CC selectively hybridize to a sequence that is at least 80% identical
CC to them. The prostate cancer-associated polynucleotide sequences
CC are differentially expressed in prostate tumour tissue or in
CC prostate cancer and are derived from the tissues of various
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC The methods of the invention are useful for diagnosing and treating
CC prostate cancer in mammals. The prostate cancer-associated genes are
CC useful for diagnosing or treating prostate cancer, as well as for
CC identifying modulators of prostate cancer or agents that inhibit
CC prostate cancer. The nucleic acid sequences are particularly useful
CC in gene therapy, as a vaccine or in antisense applications.
CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide
CC sequences.
XX
XX Sequence 3004 BP; 874 A; 679 C; 710 G; 741 T; 0 other;
SQ
Query Match 94.7%; Score 2855.4; DB 24; Length 3004;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 2976; Conservative 0; Mismatches 26; Indels 15; Gaps 9;
QY 1 CCCTATGGCGATTGGCGGCTGCAGACACGAGCTCAGTTCCTCCCTGCCCTAGTCTGAG 60
DB 1 CCCTATGGCGATTGGCGGCTGCAGACACGAGCTCAGTTCCTCCCTGCCCTAGTCTGAG 60
QY 61 CTTAGTGGTGGGACTCAGCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCT 120
DB 61 CTTAGTGGTGGGACTCAGCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCT 120
QY 121 TTCCTACACTTTTCCTGCCTAGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
DB 121 TTCCTACACTTTTCCTGCCTAGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
QY 181 ACCTGTGGGGAGGAGCGCCTGGAGGAGCTTAGAGACCCAGCCGCGGCGGTGATCTCACC 240
DB 181 ACCTGTGGGGAGGAGCGCCTGGAGGAGCTTAGAGACCCAGCCGCGGCGGTGATCTCACC 240
QY 241 ATGTGGCGATTTCGAGGCGCGCCCTGGAGCTGCTAGAGATCCGGAAGCAGACGCCCGGAG 300
DB 241 ATGTGGCGATTTCGAGGCGCGCCCTGGAGCTGCTAGAGATCCGGAAGCAGACGCCCGGAG 300
QY 301 GTCTGCGAAGCCACCAAGACTCGCGCTCTGGAGAAGCGGTGAGAGAGCGCCACCGCGG 360
DB 301 GTCTGCGAAGCCACCAAGACTCGCGCTCTGGAGAAGCGGTGAGAGAGCGCCACCGCGG 360
QY 361 TCTCCGCGCTGTCTGCACCTGTGCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 420
DB 361 TCTCCG - GCCTGTCTGCACCTGTGCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 419
QY 421 TTACCAAGTGTCTTGAATTGATAGTGGCTTCTCTTTGTCTAGTCTCATATAAGAACTACAG 480
DB 420 TTACCAAGTGTCTTGAATTGATAGTGGCTTCTCTTTGTCTAGTCTCATATAAGAACTACAG 479
QY 481 CTATCAGGAGGAGATCGCAGAGGAGGTGAGAGACACACACCATGTTCTGCAGAGCT 540
DB 480 CTATCAGGAGGAGATCGCAGAGGAGGTGAGAGACACACACCATGTTCTGCAGAGCT 539
QY 541 CAAGGATCTCAAGATCAGAGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
DB 541 CAAGGATCTCAAGATCAGAGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
```

```
DB 540 CAAGGATCTCAAGATCAGAGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
QY 601 TAACGAGTCTTTCAGAGGAGGAGCAGCAGGAGCTCAGAGAGCTGCAAGAGCAACCGCTGCCCAT 660
DB 600 TAACGAGTCTTTCAGAGGAGGAGCAGCAGGAGCTCAGAGAGCTGCAAGAGCAACCGCTGCCCAT 659
QY 661 CTGTCAGACATTTCTGAGAGAACATACAAAGAGTCTTCTCTCAAGAAAAACCAAGTCG 720
DB 660 CTGTCAGACATTTCTGAGAGAACATACAAAGAGTCTTCTCTCAAGAAAAACCAAGTCG 719
QY 721 GAGCCGAGTCTATCTTTCACACTTTGGCAGAGAGTATTGGCAACTGATTTTCCACAGTT 780
DB 720 GAGCCGAGTCTATCTTTCACACTTTGGCAGAGAGTATTGGCAACTGATTTTCCACAGTT 779
QY 781 TGAACGGCTGAATTTGTCACACTTTCAGAGAACATTTGGCAAGCAGCAAAATAAAGAAAGCAG 840
DB 780 TGAACGGCTGAATTTGTCACACTTTCAGAGAACATTTGGCAAGCAGCAAAATAAAGAAAGCAG 839
QY 841 GAATCTTTGGAAGAGAGAGACTTTGAAAAACAATTTGCAGAGCAAGCAGTTTCGAGCAGG 900
DB 840 GAATCTTTGGAAGAGAGAGACTTTGAAAAACAATTTGCAGAGCAAGCAGT---GCAGC 895
QY 901 AGTTCCAGTGGAGGTTATCAAGAAATCTCTTGTGAAGAGTCTTTTAAATATGTTACGA 960
DB 896 AGAGTCCAGTGGAGTTATCAAGAAATCTCTTGTGAAGAGTCTTTTAAATATGTTACGA 955
QY 961 GGAAGATGAACACATCTTTGGGTTGGGAGCAGCCTTTAAAGATTTTAAACAGCTT 1020
DB 956 GGAAGATGAACACATCTTTGGGTTGGGAGCAGCCTTTAAAGA-TTTTAAACAGCTT 1014
QY 1021 CAGTACCTTCTCAACAGAGCAGCAGCAGTTCGCAAGAGCAGGAAAAAGGGCAGGCTGA 1080
DB 1015 CAGTACCTTCTCAACAGAGCAGCAGCAGTTCGCAAGAGCAGGAAAAAGGGCAGGCTGA 1073
QY 1081 GGACGCTCCATCTATGCTGATGAAGAGAGTATTTTCTACATGTTTACTACTTCTT 1140
DB 1074 GGACGCTCCATCTATGCTGATGAAGAGAGTATTTTCTACATGTTTACTACTTCTT 1133
QY 1141 CCCTAAGAGAACACCTCCCTGATCTTCCCGCATCATAAAGCAGCTGCTCAGCATTT 1200
DB 1134 CCCTAAGAGAACACCTCCCTGATCTTCCCGCATCATAAAGCAGCTGCTCAGCATTT 1193
QY 1201 ATATGAACGGAAGTGGAGTGTCTTAATGCTTCCCTGCTTCCATATGATTTGCAGCGA 1260
DB 1194 ATATGAACGGAAGTGGAGTGTCTTAATGCTTCCCTGCTTCCATATGATTTGCAGCGA 1253
QY 1261 GTTTGTGAATCAGCCCTACTTGTGTACTCCGTTCACATGAAAGCAGCAGGAGCAGCT 1320
DB 1254 GTTTGTGAATCAGCCCTACTTGTGTACTCCGTTCACATGAAAGCAGCAGGAGCAGCT 1313
QY 1321 GTCCCGCAGAAACCCAGTCTCGCTGGTGGATTCCCATCATGCTATTTGCAAGACATT 1380
DB 1314 GTCCCGCAGAAACCCAGTCTCGCTGGTGGATTCCCATCATGCTATTTGCAAGACATT 1373
QY 1381 TCCATTCCATTTCATGTTTGACAAAGATATGCAATTTCTGCAATTTGGCAATGGCATCAG 1440
DB 1374 TCCATTCCATTTCATGTTTGACAAAGATATGCAATTTCTGCAATTTGGCAATGGCATCAG 1433
QY 1441 AAGGCTGATGAACAGGAGAGACTTTCAAGGAAAGCCTTAATTTTGAAGAATACTTTGAAT 1500
DB 1434 AAGGCTGATGAACAGGAGAGACTTTCAAGGAAAGCCTTAATTTT---GAATACTTTGAAT 1490
QY 1501 TCTGACTCCAAAAATCAACAGAGCTTTAGCGGATCATGACTATGTTGAATATGCAAGTT 1560
DB 1491 TCTGACTCCAAAAATCAACAGAGCTTTAGCGGATCATGACTATGTTGAATATGCAAGTT 1550
QY 1561 TGTGTACAGTGTAGAGATGGAGCAACTCTGTGAAGAAATCTTCAAGGGTTATGGACCT 1620
DB 1551 TGTGTACAGTGTAGAGATGGAGCAACTCTGTGAAGAAATCTTCAAGGGTTATGGACCT 1610
QY 1621 CAAAGGCCAAATCATCTACATTTGTAATCCAGTGGCAATCTGTTTGGGTCACCTG 1680
DB 1611 CAAAGGCCAAATCATCTACATTTGTAATCCAGTGGCAATCTGTTTGGGTCACCTG 1670
```

QY 1681 TGTGGACAGATTAGAGATTTTACAGACGAGGGCTCTACCTCTCAGACATCCCAATTCA 1740
DB 1671 TGTGGACAGATTAGAGATTTTACAGACGAGGGCTCTACCTCTCAGACATCCCAATTCA 1730
QY 1741 CAATGCACTGAGGATGTGTCTTAATAGGGGAACAGCCGAGCTCAAGATGCCGTGA 1800
DB 1731 CAATGCACTGAGGATGTGTCTTAATAGGGGAACAGCCGAGCTCAAGATGCCGTGA 1790
QY 1801 GAAGAGCTGGGAAGCTGAAGCTACCTTTGAGCAAGCCACCAAGCCCTGGAGGAGGA 1860
DB 1791 GAAGAGCTGGGAAGCTGAAGCTACCTTTGAGCAAGCCACCAAGCCCTGGAGGAGGA 1850
QY 1861 GAAGAAAGACAGTAGACCTTCTGTCTCCATATTTCCCTGTGAGGTTCTCAGCAGCT 1920
DB 1851 GAAGAAAGACAGTAGACCTTCTGTCTCCATATTTCCCTGTGAGGTTCTCAGCAGCT 1910
QY 1921 GTGCAAGGCAAGTTGTGCAAGCAAGAGTTCAAGTATGTCAACATGCTCTCTCAGA 1980
DB 1911 GTGCAAGGCAAGTTGTGCAAGCAAGAGTTCAAGTATGTCAACATGCTCTCTCAGA 1970
QY 1981 CATCGTTGGGTTCACTGCCATCTGCTCCAGTGTCTACCGCTGCAGGTCAATCACCATGCT 2040
DB 1971 CATCGTTGGGTTCACTGCCATCTGCTCCAGTGTCTACCGCTGCAGGTCAATCACCATGCT 2030
QY 2041 CAATGCACTGTACACTGCTTGCAGCAGCAGTGTGGAGAGCTGGATGCTTACAAGTGA 2100
DB 2031 CAATGCACTGTACACTGCTTGCAGCAGCAGTGTGGAGAGCTGGATGCTTACAAGTGA 2090
QY 2101 GACCATTTGGGATGATGCTTGTGTAGC - TGGGGGATTTACAAAGAGTGTACTCATG 2159
DB 2091 GACCATTTGGGATGATGCTTGTGTAGC - TGGGGGATTTACAAAGAGTGTACTCATG 2149
QY 2160 CTGTTCAGATAGGCTGATGCCCTGAAGATGATGAGGCTCTCTGATGAAGTATGCTC 2219
DB 2150 CTGTTCAGATAGGCTGATGCCCTGAAGATGATGAGGCTCTCTGATGAAGTATGCTC 2209
QY 2220 CCATGAGAACCTATCAAGATGGAAATGGACTGCACTGTGATCAGTATTTTGTGGCG 2279
DB 2210 CCATGAGAACCTATCAAGATGGAAATGGACTGCACTGTGATCAGTATTTTGTGGCG 2269
QY 2280 TCGTTGGATTAATAATCCCGCTTACTGCTTTTGGAAACAATGTCACTCTCGCTAACA 2339
DB 2270 TCGTTGGATTAATAATCCCGCTTACTGCTTTTGGAAACAATGTCACTCTCGCTAACA 2329
QY 2340 AATTGAGTCTGCACTGATACCAAGAAATCAATGTGAGCCCAACAACTTACAGATTAC 2399
DB 2330 AATTGAGTCTGCACTGATACCAAGAAATCAATGTGAGCCCAACAACTTACAGATTAC 2389
QY 2400 TCAAGACTGCTGCTGTTGCTGTTTACCCCTCGATCAAGGGAGAACTTCCACCAACT 2459
DB 2390 TCAAGACTGCTGCTGTTGCTGTTTACCCCTCGATCAAGGGAGAACTTCCACCAACT 2449
QY 2460 TCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGCTTACCAAGAGGAACAACT 2519
DB 2450 TCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGCTTACCAAGAGGAACAACT 2509
QY 2520 CAAACCACTCTTCCAAAGAAAGATGTGAAGATGGCAATGCAATTTTTTAGGCAAG 2579
DB 2510 CAAACCACTCTTCCAAAGAAAGATGTGAAGATGGCAATTTTTTAGGCAAG 2567
QY 2580 CATCAGAAATAGATTAGCAACCTATATACCTATTTTAAAGTCTTTGGGGTTTCACTATT 2639
DB 2568 CATCAGAAATAGATTAGCAACCTATATACCTATTTTAAAGTCTTTGGGGTTTCACTATT 2627
QY 2640 GAAGATGTGTAGAGCTCTCAAGAGCACTTTAGGATTTGTAGATGGCTTAACAAGCAGTATT 2699
DB 2628 GAAGATGTGTAGAGCTCTCAAGAGCACTTTAGGATTTGTAGATGGCTTAACAAGCAGTATT 2687
QY 2700 AAAATTTAGAGCCCAAGTCAATCTTTCTCTGTTTAAATGACAAAATGTACTCACT 2759
DB 2688 AAAATTTAGAGCCCAAGTCAATCTTTCTCTGTTTAAATGACAAAATGTACTCACT 2747

QY 2760 TCAGTACTTCAGCTCTTCAAGAAAAAACCCTTAAAAAGCTACTTTTGTGGGAGT 2819
DB 2748 TCAGTACTTCAGCTCTTCAAGAAAAAACCCTTAAAAAGCTACTTTTGTGGGAGT 2807
QY 2820 ATTTCTATTATTAACACGACCTTACTACCTGTACTGCTCAAAATTCAGCACCTTGTACATAT 2879
DB 2808 ATTTCTATTATTAACACGACCTTACTACCTGTACTGCTCAAAATTCAGCACCTTGTACATAT 2867
QY 2880 ATCAGATAATTTAGTCAATTTGTAACAACTGTAGGAGTCACTGCAATCTCATATCTCTGG 2939
DB 2868 ATCAGATAATTTAGTCAATTTGTAACAACTGTAGGAGTCACTGCAATCTCATATCTCTGG 2927
QY 2940 TGAATGCCATGTTTATTAAGTGTGTTTGTGATAG - TGTGCTCAAAAAAAGAAAAA 2998
DB 2928 TGAATGCCATGTTTATTAAGTGTGTTTGTGATAGTGTGCTCAAAAAAAGAAAAA 2987
QY 2999 AAAAAAAGAAAAA 3015
DB 2988 AAAAAAAGAAAAA 3004

RESULT 8
AAZ51684
ID AAZ51684 standard; cDNA; 2715 BP.
XX
AC AAZ51684;
XX AC
DT 04-JUL-2000 (first entry)
XX
DE Human cyclic nucleotide-associated protein-3 (CNAP-3) cDNA.
XX
KW Cyclic nucleotide-associated protein-3; CNAP-3; human; cytostatic;
KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;
KW immunomodulatory; anti-asthmatic; anti-anaemic; anti-diabetic; diagnosis;
KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;
KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;
KW anti-infertility; anti-allergic; vasotropic; immunosuppressive;
KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;
KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;
KW neurological; vision; reproductive; smooth muscle; ss.
XX
OS Homo sapiens.
XX
FH Key
CDS 440..2512
FT FT /*tag= a
FT FT /product= "Human CNAP-3 protein"
FT FT /note= "Shares 89% identity to human soluble guanylate
FT FT cyclase large subunit"
FT FT /tag= b
FT FT 863..892
FT FT /bound_molety= "Primer or Probe"
FT FT /note= "Useful for amplification or hybridisation
FT FT techniques"
XX
PN WO200014248-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20287.
XX
PR 04-SEP-1998; 98US-0148904.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;
XX WPI; 2000-256994/22.
XX P-PSDB; AAY70475.
XX
PT Isolated cyclic nucleotide associated proteins useful for preventing,
PT diagnosing and treating cell proliferative, autoimmune/inflammatory,
PT neurological, vision, reproductive and smooth muscle disorders -

XX

PS Example 3; Page 71-72; 78pp; English.

XX The present sequence is the cDNA encoding human cyclic nucleotide
CC associated protein-3 (CNAP-3), identified in Incyte clone 159278,
CC that is isolated from ADENINB01 cDNA library. It is expressed in
CC nervous, reproductive, cardiovascular and developmental tissues.
CC CNAP sequences may be used for prevention, treatment and diagnosis of
CC diseases associated with altered CNAP expression such as, cell
CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukemia,
CC lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/
CC inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple
CC sclerosis and psoriasis), neurological disorders (e.g. epilepsy,
CC Alzheimer's/Parkinson's disease and strokes), vision disorders (e.g.
CC conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),
CC reproductive disorders (e.g. infertility, uterine fibroids, ectopic
CC pregnancies and impotence) and smooth muscle disorders (e.g. angina,
CC anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also
CC be used for gene therapy.

XX Sequence 2715 BP; 799 A; 604 C; 662 G; 650 T; 0 other;

Query Match 78.5%; Score 2368; DB 21; Length 2715;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2382; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY	412	ACATCCCACTACCAAGTCTCCTTGAATGATAGTGGCTTCTGTTGTCTAGTCTCATATAA	471
DB	328	ACATCCCACTACCAAGTCTCCTTGAATGATAGTGGCTTCTGTTGTCTAGTCTCATATAA	387
QY	472	GAACATACCTCATCAGGAGGATCGCAGCAGGATGAGAGACACCAACCATGTTCTG	531
DB	388	GAACATACCTCATCAGGAGGATCGCAGCAGGATGAGAGACACCAACCATGTTCTG	447
QY	532	CACGAAGCTCAAGGATCTCAAGATCACAGGAGTGTCTTCTCTTACTGSCACCG	591
DB	448	CACGAAGCTCAAGGATCTCAAGATCACAGGAGTGTCTTCTCTTACTGSCACCG	507
QY	592	TCAAGTTCCTAACGAGTCTTACAGAGGAGCAGCAGGAGCTCAGAGAGTGCACCAAGCAAC	651
DB	508	TCAAGTTCCTAACGAGTCTTACAGAGGAGCAGCAGGAGCTCAGAGAGTGCACCAAGCAAC	567
QY	652	CGTGCCCATCTCAGAGATCTCCTCAGAGACATACAGAGAGTCTTCTCTCCTTACTGSCACCG	711
DB	568	CGTGCCCATCTCAGAGATCTCCTCAGAGACATACAGAGAGTCTTCTCTCCTTACTGSCACCG	627
QY	712	AACCACTCGAGCGAGTCTATCTTACACTTTGGCAGAGATATTTGCAAACTGATTTT	771
DB	628	AACCACTCGAGCGAGTCTATCTTACACTTTGGCAGAGATATTTGCAAACTGATTTT	687
QY	772	CCGAGAGTTGAACGGCTGAATGTTGCACCTCAGAGAACATTTGGCAAAACACAAATAAA	831
DB	688	CCGAGAGTTGAACGGCTGAATGTTGCACCTCAGAGAACATTTGGCAAAACACAAATAAA	747
QY	832	AGAAAGCAGGAATCTTTCGAAGAGAGACATTTGAAAACAAATTCGACAGCAAGCAGT	891
DB	748	AGAAAGCAGGAATCTTTCGAAGAGAGACATTTGAAAACAAATTCGACAGCAAGCAGT	807
QY	892	TGCAGCAGGAGTTCAGTGGAGTTATCAAGAAATCTTGTGTGAAGAGGTTTTTAAAT	951
DB	808	TGCAGCAGGAGTTCAGTGGAGTTATCAAGAAATCTTGTGTGAAGAGGTTTTTAAAT	867
QY	952	ATGTTACAGGAGAGTGAATAACATCCTTGGGGTGTGGAGCACCCTTAAAGATTTTTT	1011
DB	868	ATGTTACAGGAGAGTGAATAACATCCTTGGGGTGTGGAGCACCCTTAAAGATTTTTT	927
QY	1012	AAACAGCTTCAGTACCCTTCTGAAACAGAGCAGCCATTGCCAAGAGCAGGAAAAGGGG	1071
DB	928	AAACAGCTTCAGTACCCTTCTGAAACAGAGCAGCCATTGCCAAGAGCAGGAAAAGGGG	987
QY	1072	CAGGCTTGAGGAGCGCTCCATCTATGCTTGGATAGGAGGATGATTTTCTACATGTTTA	1131
DB	988	CAGGCTTGAGGAGCGCTCCATCTATGCTTGGATAGGAGGATGATTTTCTACATGTTTA	1047

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

Db 269 GAACCTACAGCTCATCAGGAGGAGATCGCAAGCAGGGTAAGAGACACCAACCACTGTTCT 328
QY 531 GCACGAAGCTCAAGGATCTCAAGATCACAGGAGAGTGTCCTTTCTCTTACTGGCACCAG 590
Db 329 GCACGAAGCTCAAGGATCTCAAGATCACAGGAGAGTGTCCTTTCTCTTACTGGCACCAG 388
QY 591 GTCAGTTCCTAA-CGAGTCTTCAGAGGAGGAGCAGGAAGCTCAGAGAGC-TGCAAGC 648
Db 389 GTCAGTTCCTAAACCGAGTCTTCANAGGAGGAGCAGGAAGCTCAGAGAGCTTGCAAGC 448
QY 649 AACCTGCCCATCTGTCACACATCTCTGAGAGAGACATACAGAAAGTCTTCTCAAG 708
Db 449 AACCTGCCCATCTGTCACACATCTCTGAGAGAGACATACAGAAAGTCTTCTTAAG 508
QY 709 AAAAAACAGT-CGAGGCGGAGTCTATCTTCCACACTTTGGCA-GAGAGTATTGCAACTG 766
Db 509 AAAAAACAGTCCGAGCGGAGTCTATCTTTACACTTTGGCANGAAGTATTTNCAACTG 568
QY 767 ATTTTCCAGAG-TTTGAACGGCTGAA-TGTTGCACTTCAGAGAA 809
Db 569 ATTTTCCAGAGTTTGAACGGCTGAATTGTTACTTCACTTCANAAAA 613

RESULT 15

ABL04726

ID ABL04726 standard; cDNA; 3094 BP.

XX AC ABL04726;

XX XX

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 8660.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX W0200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB60623.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 8660; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3094 BP; 782 A; 727 C; 730 G; 855 T; 0 other;
Query Match 8.7%; Score 260.8; DB 23; Length 3094;
Best Local Similarity 61.2%; Pred. No. 2e-58;
Matches 421; Conservative 0; Mismatches 267; Indels 0; Gaps 0;
QY 1727 GACATCCCAATTCACAATGCACCTGAGGAGTGTGCTCTTAATAGGGGAAACAGCCGAGCT 1786
Db 2 GACATACCGTGCACGACGCCACCGAGGAGTGATCTCTGTGGGCGAAGAGCTCGGGCC 61
QY 1787 CAAGATGCCCTGAAGAGAGGCTGGGGAAGCTGAAGCTACCTTTGAGCAGCCCAACAA 1846
Db 62 CAGATGGACTGCCCGCGCATGGACAAGATCAAGAACAGCATTTGAGGAGGCCACATCG 121
QY 1847 GCCCTGGAGGAGGAGAGAAAAGACAGATGACCTTCTGTCTCCATATTTCCCTGTGAG 1906
Db 122 GCGGTGACCAAGGAGGCGGAAGAACGTCAGTCTGTGCACTCATCTTTCCCGCGGAG 181
QY 1907 GTTGCTCAGCAGCTGTGGCAAGGCAAGTTGTGCAAGCCCAAGAGTTTCAGTAATGTCA 1966
Db 182 ATCGCCGAGAAAGCTTTGGCTGCTCCATTGACGCCAAGACGATATCCCGATGTTACG 241
QY 1967 ATGCTCTTCTCAGACATCGTTGGTTCACCTGCCATCTGCTCCAGTGTCTACCCCTGCAG 2026
Db 242 ATCCTGTTCAGCGACATCGTTGGCTTCACTAGATCTGCTCGCGGCGCCTCCCTCATG 301
QY 2027 GTATCACCATGCTCAATGCACCTGACACTGCTTCGACACAGTGTGGAGAGCTGGAT 2086
Db 302 GTGATCAGCATGCTGGAGGGGCTCTACAAGGACTTCGACGAGTTCTGCGACTTCTTCGAC 361
QY 2087 GTCTACAAGTGGAGACCATTTGGGATGCCCTATTTGTGTAGCTGGGGGATTCACAAAGAG 2146
Db 362 GTGTACAAGTGGAGACCATTCGGGGATCCCTACTGCTGCGGCGCAGTGGACTCCACCGAG 421
QY 2147 AGTGATACTCATGCTGTTTCAGATAGCGCTGATGGCCCTGAAGATGATGGAGCTCTCTGAT 2206
Db 422 TCCATCTACGACGCCACAGGTTGCTGATGGCCCTGAAGATGATGCGAGCTGCTGTCG 481
QY 2207 GAAGTTATGTCCTCCCATGGAGAACCTTATCAAGATGCGAATTTGAGCTGCACTCTGGATCA 2266
Db 482 AAGCACATTATCCACACGCGGAGCAAAATTAATGAGGATCGGCTGCACACGGGCACT 541
QY 2267 GTTTTGTGCTGCGCTGTTGGAGTTAAATGCCCCGTTACTGCTCTTTTGGAAACAATGTC 2326
Db 542 GTCTTGGCGGGAGTGGTGGCGAGGAAGATGCCAGGATTTGCCCTTTTCGGGCACAGCGTC 601
QY 2327 ACTCTGGCTAACAAATTTGAGTCTCCTGAGTGTACACGAGAAATCAATGTGAGCCCAACA 2386
Db 602 ACCATCCCAACAGTTTCGAGTCGGGCGAGCGGCTGAGATTAATGTTAGCCCAACC 661
QY 2387 ACTTACAGATTACTCAAAAGACTGTCCTG 2414
Db 662 ACCAAGGAGTAAGTGGTACTCTACTG 689

Search completed: July 1, 2003, 13:26:14

Job time : 441.134 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 1, 2003, 12:30:42 ; Search time 2760.9 Seconds
(without alignments)
17686.068 Million cell updates/sec

Title: US-09-762-767A-1
Perfect score: 3015
Sequence: 1 cccctatggcgattggcggg.....aaaaaaaaaaaaaaaaaaaaa 3015

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba.*
2: em_esthum.*
3: em_estmu.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1609.8	53.4	3054	11 AK004815	AK004815 Mus muscu
2	971.2	32.2	1169	13 BM544920	BM544920 AGENCOURT
3	882.4	29.3	884	14 BQ707076	BQ707076 AGENCOURT
4	738.2	24.5	850	9 AL543909	AL543909 AL543909
5	720.8	23.9	777	12 BG220276	BG220276 RST40049
6	707.4	23.5	829	9 AL563000	AL563000 AL563000

C	7	707	23.4	852	9	AL570204	AL570204
C	8	692.4	23.0	803	9	AI719730	AI719730
C	9	692.2	23.0	708	9	AA555318	AA555318
C	10	688.2	22.8	741	14	BQ74520	UI-H-E21-
C	11	687	22.8	1288	11	BC012627	Homo_sapi
C	12	673.4	22.3	824	13	BI771039	603055368
C	13	670	22.2	806	12	BG743485	602635511
C	14	655.2	21.7	767	9	AL527890	AL527890
C	15	654.8	21.5	658	14	BQ549434	1106h09.x
C	16	649.4	21.5	653	13	BM665903	UI-B-DX1-
C	17	625.2	20.7	1183	14	BM925915	AGENCOURT
C	18	624	20.7	777	12	BG203016	RST22385
C	19	620.2	20.6	716	9	AA779213	zj53c05.s
C	20	615	20.4	793	13	BI551700	603197377
C	21	605.4	20.1	964	13	BM469193	AGENCOURT
C	22	593	19.7	600	14	BQ581283	1106h09.y
C	23	567.4	18.8	595	10	AW513686	XB87F09.x
C	24	554.4	18.4	824	13	BI769310	603054578
C	25	551.8	18.3	644	10	AW006446	wt16903.x
C	26	531.4	17.6	638	12	BG169925	602321619
C	27	506.2	16.8	631	10	AW951558	AW951558
C	28	505.4	16.8	539	9	AA189150	zq44g01.s
C	29	504.8	16.7	530	9	AL120503	DKFZP761N
C	30	495.2	16.4	734	13	BI551777	603197469
C	31	486.4	16.1	583	10	AW973582	EST385682
C	32	482.8	16.0	487	9	AL123854	qg74f08.x
C	33	481	16.0	497	9	AL120382	DKFZP761D
C	34	478	15.9	1068	12	BG294057	602390917
C	35	477.4	15.8	583	9	AA193614	zr42g01.f
C	36	449.4	14.9	482	9	AA988462	or77n08.f
C	37	448.6	14.9	727	13	BI730880	603350511
C	38	442.2	14.7	705	9	AL598467	DKFZP3131
C	39	436.4	14.5	720	13	BI602173	603246294
C	40	435	14.4	911	12	BF679959	602154663
C	41	427.6	14.2	746	9	AL601127	DKFZP313N
C	42	424.8	14.1	447	13	BM662859	UI-B-CKO-
C	43	419	13.9	533	13	BM146581	TCAAP1E2
C	44	406.8	13.5	446	9	AA259214	nc17el0.f
C	45	388.6	12.9	675	13	BI546815	603189713

ALIGNMENTS

RESULT 1	AK004815	AK004815	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
LOCUS	AK004815	AK004815	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
DEFINITION	AK004815	AK004815	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
ACCESSION	AK004815	AK004815	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
VERSION	AK004815.1	AK004815.1	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
KEYWORDS	HTC; CAP trapper.	HTC; CAP trapper.	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
SOURCE	Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA, clone:1200016007.	Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA, clone:1200016007.	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
ORGANISM	Mus musculus	Mus musculus	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
AUTHORS	Carninci, P. and Hayashizaki, Y.	Carninci, P. and Hayashizaki, Y.	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
TITLE	High-efficiency full-length cDNA cloning	High-efficiency full-length cDNA cloning	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	Meth. Enzymol. 303, 19-44 (1999)	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
MEDLINE	99279253	99279253	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
PUBMED	10349636	10349636	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
REFERENCE	Carninci, P. and Hayashizaki, Y.	Carninci, P. and Hayashizaki, Y.	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
AUTHORS	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	Genome Res. 10 (10), 1617-1630 (2000)	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
MEDLINE	20499374	20499374	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002
PUBMED	11042159	11042159	3054 bp	mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200016007:guanylate cyclase 1, soluble, alpha 3, full insert sequence.	HTC 19-JAN-2002

Db	441	GCACAGAGCTACCTGCGACACCCCTGGCAGAGAGTATTTTGGCAAGCTCATCTTCCCGAGGT	500
Qy	780	TTGAACGGCTGAATCTTGACATTCAGAGAAACATTTGGCAAGCACAAAATAAAGAAAGCA	839
Db	501	GTGAGCGACTGAACCTTGACATTCAGAGAACCTTTGGCAAGCATATAATAGAGAAACA	560
Qy	840	GGAAATCTTTGGAAGAGAGAACTTTTGAAAAAACAATTTGCAGAGCAAGCAGTTGCGACAG	899
Db	561	GGAAATCTTTAGAAAAGAGAACCTTTGAAAAAATAATCGCAGAAAGCAATTTGCAGCAG	620
Qy	900	GAGTTCACGTGAGGGTTATCAAGAATCTCTTGGTGAAGAGGTTTAAAAATATGTTACG	959
Db	621	GTCCCCAGTGGAGGGCTTCAAGAGCTCTTGGGGGAGGAGCTGTTCAGATCTTGCCTATG	680
Qy	960	AGGAAGATGAAACATCCCTTTGGGGTGGTGGAGGCACCCCTTAAAGATTTTAAAAACAGCT	1019
Db	681	AGGAAGATGAGCACATTTTGGGCGTGTGGCGGCACCCCTGAGGACTTCTTAATAGCT	740
Qy	1020	TCAGTACCCCTTCGAAACAGAGCAGCCATTTGCCAAGAAGAGAAAGAGGGCAGGGTTG	1079
Db	741	TCAGCAGCGTCTTCAGCAGAGAGCAGCCACTGTCCAAAGAGCGGAGGGCGGAGCACTGG	800
Qy	1080	AGGACCCCTCCACTTCATGCCCTGGATAGAGGAGATGATTTTCTACATGTTTACTTCTCT	1139
Db	801	AGATGCCCTCCACTTCATGCCCTGGACAGGACAGGACTTCTTAATGTTTACTACTCT	860
Qy	1140	TCCTTAAGAGAACCCCTCCCTGATCTTCCCGGCATCATAAAGCAGCTCTCACGCTAT	1199
Db	861	TCGCCGAAGAGAACCCAGCCCTTCTCCCTGGTATCATTAAGCGGCTGCTCGCATAC	920
Qy	1200	TATATGAACGGAAGTGGTGGTGTGTTAATGCCCTCCCTGCTTCATTAATGATTCGACGG	1259
Db	921	TGTACGAAAGCCAGCTGGAGGTGTCCCTGATGCCCTCCCTGCTTCCGAAAGTCACTGTACCG	980
Qy	1260	AGTTTCTGATCAGCCCTACTTGTGTACTCGCTTCACATGAAGAGCACCAGCCATCCC	1319
Db	981	AGTTTGTGAACAGCCCTATTGTGCTACTCCGTTTCATGTGAAGAGCACCAGCCGCTCC	1040
Qy	1320	TGTCCCCCGAGCAAAACCCAGTCCCTCGCTGATTCACACATCGCTATTTCGCAAGACAT	1379
Db	1041	TGTCCCCCGAGCAAGCCAGTCCCTCGCTAGTATCCCGCTTCGCTCTCTTCGAAAGACTT	1100
Qy	1380	TTCCATTCATTTGATTTGACAAGATATGACAATTTGCAATTTGGCAATGGCATCA	1439
Db	1101	TCCCGTTCCATTTGCTTGGACCGAGACCTGGCCATCCTCGAGCTGGGTAAACGGCATCA	1160
Qy	1440	GAGGCTGATGACAGGAGAGACTTTCAGGAAAGCCCTAAATTTCAAGAACTATTGAAA	1499
Db	1161	GAGGCTGTGTAACAAGAGGACTTCCAGGGAAGCCCACTTTGAAGAGTCTTTGAAA	1220
Qy	1500	TTCTGACTCCAAAAATCAACAGAGAGTTTGGGGGATCATGACTATGTTGAATGCGAT	1559
Db	1221	TTCTAACTCCCAAAATCAACAGACATTTAGTGGCATCATGACAATGTTGAACATGCAGT	1280
Qy	1560	TTGTTGTAGAGTGGAGATGGGACAACCTTGTGAAGAAATCTTCAAGGTTTATGGACC	1619
Db	1281	TTGTCTATCCGGGTGAGGAGATGGGATAACTCGGTGAAGAAATCGTCAAGGGTTATGGATC	1340
Qy	1620	TCAAGGCCAAATGATCTACATTTGTTGAATCCAGTGCATCTTGTGTTGGGGTCACCCCT	1679
Db	1341	TCAAGGTCAAATGATCTACATCGTTGAATCCAGTGCCATCTTGTCTTGAAGGTCACCAT	1400
Qy	1680	GTCTGACAGATTTAGAAGATTTTACAGGACGAGGCTCTACTCTTCAGACATCCCAATTC	1739
Db	1401	GTCTGACAGGCTGGAGATTTTACAGAGCGGGGCTCTATCTGTCCACATCCCAATTC	1460
Qy	1740	ACAATGCACGTAGGGATGTGCTTTAATAGGGGAACAAGCCCGAGCTCAAGATGGCCCTGA	1799
Db	1461	ATAACGCCCTGAGGGATGTGCTTGTATAGGGGAGCAGCAGCGGCTCAAGATGGCCCTCA	1520
Qy	1800	AGAGAGGCTGGGAGCTTGAAGCTTACCTTTGACGAAGCCCAACCCCTCGAGGAGG	1859
Db	1521	AGAGAGGTTGGGAGAGCTGAAGGCAACCCCTGGAGCATGCCACACAGCCCTGGAGGAGG	1580
Qy	1860	AGAAGAAAAGACAGTACAGCTTCTGCTCCATATTTCCCTGTGAGGTGCTCAGCAGC	1919
Db	1581	AGAAGAGAGGACAGTGGATCTGCTGCTCTATCTTCCCTCTGAGGTGCTCAGCAGC	1640
Qy	1920	TGTGGCAAGGGCAAGTGTGTGCAAGCCAAAGATTGAGTAATGTACCATGCTCTTCTCAG	1979
Db	1641	TGTGGCAAGGACAAAATTTGCAAGCCAAAGAAATTCAGCGAGGTCAACATGCTTTCTCAG	1700
Qy	1980	ACATCGTTGGGTTCACTGCCATCTGCTCCCATGCTCACCGCTCGAGGTCAATCAGTGC	2039
Db	1701	ATATCGTAGGGTTCACTGCTATCTGCTCTCACTGTTCACCTCTCGAGGTCAATCAGTGC	1760
Qy	2040	TCAATGCACGTGTACACTCGCTTCGACACGACAGTGTGGAGAGCTGGATCTTACAAAGTGG	2099
Db	1761	TCAACGCTCTCTACACTCGCTTTGACACGACAGTGTGGAGAGCTGGATCTTACAAAGTGG	1820
Qy	2100	AGACATNTGGCGATGCTATTGTGTAGCTGGGGGATTAACAAGAGAGAGTGAATCATG	2159
Db	1821	AGACCATCGGGATGCATATTGTGTGGCAGGTGGATTCACAGAGAGAGTCAACCCATG	1880
Qy	2160	CTGTTTCAGATACGCTGATGGCCCTGAAGATGATGGAGCTCTCTGATGAAGTATGTCTC	2219
Db	1881	CTGTCCAGATACACTGATGGCCCTGAAGATGATGGAGCTCTCCAATGAGGTCACTGCTC	1940
Qy	2220	CCATGGAGAACCTATCAAGATGCGAATTTGGACTGCATCTCTGGATCAGTCTTTGCTGGCG	2279
Db	1941	CCCACGGAGAACCTATCAAGATGCGAATTTGGACTACATCTTGGATCAGTCTTTGCTGGAG	2000
Qy	2280	TGTTGGAGTTTAAATGCCCCCTTACTGTCTTTTTTGGAAACAATGTCTACTCTGCTAACA	2339
Db	2001	TTGTCGGAGTGAAGATGCCCGGTATTGCTGTTTGGAAACAATGTCTACTCTGGCTAACA	2060
Qy	2340	AATTTGAGTCTTGCAGTGTACCACGAAAATAAATCAATGTCCAGCCCAACACTTACAGATTAC	2399
Db	2061	AATTTGAATCTTGCAGTGTGCTCGGAAATCAATGTCCAGCCCAACACTACAGAGTTAC	2120
Qy	2400	TCAAAGACTGTCTGCTTTCGTTTACCCCTCGGATCAAGGAGGAGAACTTCCACCAAACT	2459
Db	2121	TCAAAGACTGTCTGCTTTCGTTTACCOCGAGATCAAGGAGGAGCTTCCACCAAACT	2180
Qy	2460	TCCCTAGTGAATCCCCGGAATCTGCCATTTTCTGGATGCTT---ACCAACAAGGAACA	2516
Db	2181	TCCCTAGTGACATTTCTGGGATCTGTCACTTTCTGGATGCTTATCACCATCAAGGACCTA	2240
Qy	2517	ACTCAAAACCATGCTTCCAAAAGAAAGATGTGGAAGATGGCAATGCCAATTTTATAGGCA	2576
Db	2241	ATTCCAAAACCATGGTTCAGGATAAGATGTGGAAGATGGAAACGCCAACCTTCTTAGGCA	2300
Qy	2577	AAGCATCAGGAATAGATTAGCAACCTATATACCTATTATAAGTCTTTTGGGTTTGATCTC	2636
Db	2301	AAGCGTCAGGATAGATTAGTCAGCCACAT-----GCTCTTATGTTTGTATGCTCC	2348
Qy	2637	ATTGAAGATGTGTAGAGCTCTGAAAGCACTTTTAGGATTTGTAGATGCTTAACAAAGCAGT	2696
Db	2349	TT	

QY 2920 CCTGCAATCTCATATCCCTGGTGGGAATGCCATGGTATTATAAG-----TGTGTTTGTGATAG 2975
 Db 2645 TCACTATGGCTGTGGCCACACAAAGCTGTGGTATTATAAGTGTATGTGTGTGATAG 2704
 QY 2976 TGTGCTCAAAAAAATAAATAAATAAATAAATAAATAA 3008
 Db 2705 TGTGCGCTGTAATAATCTCTGAATAGAAACAACA 2737

RESULT 2

BM544920

LOCUS BM544920 1169 bp mRNA linear EST 20-FEB-2002
 DEFINITION AGENCOURT_6497474 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5588694
 5', mRNA sequence.

ACCESSION

BM544920

VERSION

BM544920.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1169)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapb-r@mail.nih.gov

Tissue Procurement: Invitrogen

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone Distribution: Agencourt Bioscience Corporation

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM12360

row: d

column: 07

High quality sequence start: 39

High quality sequence stop: 705.

Location/Qualifiers

1..1169

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5588694"

/clone_lib="NIH_MGC_125"

/lab_host="DH10B"

/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;

Site 1: ECoRV (destroyed); Site 2: NotI; RNA source pool

of three ovaries, from females ranging in age from 38 to

49 yo. Library is oligo-dT primed and directionally cloned

(ECoRV site is destroyed upon cloning). Average insert

size 2.1 kb, insert size range 1-3.5 kb. Library is

normalized and enriched for full-length clones and was

constructed by C. Gruber (Invitrogen). Research Genetics

tracking code 036."

BASE COUNT 350 a 250 c 275 g 292 t 2 others

ORIGIN

Query Match

Best Local Similarity 32.2%; Score 971.2; DB 13; Length 1169;

Matches 1074; Conservative 0; Mismatches 44; Indels 7; Gaps 7;

QY 812 TTGCAAGACACAAATAAAGAAAGACAGGAATCTTTGGAAGAGAGACTTTGAAAAA 871

Db 48 TTGCAAGACACAAATAAAGAAAGACAGGAATCTTTGGAAGAGAGACTTTGAAAAA 107

QY 872 ACAATTCAGAGCAGTGTGACGAGGAGTTCACGTGGAGGTTATCAAGATCTCTT 931

Db 108 ACAATTCAGAGCAGTGTGACGAGGAGTTCACGTGGAGGTTATCAAGATCTCTT 167

QY 932 GGTGAAGAGGTTTTAAATATGTTACGAGGAGATGAAACATCTTTGGGGTGGTGA 991

Db 168 GGTGAAGAGGTTTTAAATATGTTACGAGGAGATGAAACATCTTTGGGGTGGTGA 227

QY 992 GGCACCTTAAAGATTTTTTAAACAGCTTCAGTACCCTTCGTAAGAACAGACGCCATTC 1051

Db 228 GGCACCTTAAAGATTTTTTAAACAGCTTCAGTACCCTTCGTAAGAACAGACGCCATTC 287

QY 1052 CAAGAAGCAGAAAAAGGGCGAGCTTGAGGACGCCCTCCATCTTATGCCTGGATAAGAG 1111

Db 288 CAAGAAGCAGAAAAAGGGCGAGCTTGAGGACGCCCTCCATCTTATGCCTGGATAAGAG 347

QY 1112 GATGATTTTCTACATGTTTACTACTTCTTCCCTTAAGAGAACACCTCCCTCAATCTTCTCC 1171

Db 348 GATGATTTTCTACATGTTTACTACTTCTTCCCTTAAGAGAACACCTCCCTCAATCTTCTCC 407

QY 1172 GGCATCATAAAGGAGCTGCTCAGTATTATATATAAAGCGAAGTGGAGTGTCTGTTAATG 1231

Db 408 GGCATCATAAAGGAGCTGCTCAGTATTATATATAAAGCGAAGTGGAGTGTCTGTTAATG 467

QY 1232 CCTCCCTGCTCCATTAATGATTGAGGAGGAGTTTGTGAATCAGCCCTACTTGTGTACTTCC 1291

Db 468 CCTCCCTGCTCCATTAATGATTGAGGAGGAGTTTGTGAATCAGCCCTACTTGTGTACTTCC 527

QY 1292 GTTCACATGAAAGCAGCAAGCCCATCCTGTCCCCAGCAAAACCCAGTCTCCTCGTGTG 1351

Db 528 GTTCACATGAAAGCAGCAAGCCCATCCTGTCCCCAGCAAAACCCAGTCTCCTCGTGTG 587

QY 1352 ATTCACATGCTGCTTATCTGCAAGACATTTCCATTTCCATTTCCATTTCCATTTCCATTTCC 1411

Db 588 ATTCACATGCTGCTTATCTGCAAGACATTTCCATTTCCATTTCCATTTCCATTTCCATTTCC 647

QY 1412 ACATTTCTGCAATTTGGCAATGGCATGAGAGGCTGATGACAGGAGAGACTTTTCAGGA 1471

Db 648 ACATTTCTGCAATTTGGCAATGGCATGAGAGGCTGATGACAGGAGAGACTTTTCAGGA 707

QY 1472 AAGCTTAATTTGAAGAATACTTTGAAATCTGACTCCAAATAACAACAGACGCTTTAGC 1531

Db 708 AAGCTTAATTTGAAGAATACTTTGAAATCTGACTCCAAATAACAACAGACGCTTTAGC 767

QY 1532 GGGATCATGACTATGTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1591

Db 768 GGGATCATGACTATGTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 827

QY 1592 GTGAAGAAATCTTCAAGGGTTATGACCTCAAGGCCAAATGATCTACATTTGTTGAATCC 1651

Db 828 GTGAAGAAATCTTCAAGGGTTATGACCTCAAGGCCAAATGATCTACATTTGTTGAATCC 887

QY 1652 AGTCAATC-TGTGTTTGGGGTACCTGTGTGACAGATTTAGAAGA-TTTTACAGGAC 1709

Db 888 AGTCAATC-TGTGTTTGGGGTACCTGTGTGACAGATTTAGAAGA-TTTTACAGGAC 947

QY 1710 GAGGGCT-CTACCTCTCAGACATCCCAATTCACATGCACTGAGGGATGTTGTTAATA 1768

Db 948 GAGGGCTCTTACCTCTCAGAAATCCCAATTC-CCATGCACTGAAAGGATGTTGTTAATA 1006

QY 1769 GGGGAACAAGCCGAGCTCAAGATGGCTGAAGAAGAGGCTGGGGAAGCTGGAAGGCTACC 1828

Db 1007 GGGGAACAAGCCGAGCTTAAAT-GCCTGAAAAAAGCTTGGGAACCTTAAGG-TACC 1064

QY 1829 CTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888

Db 1065 CTTGAGCAAGCC-CCAGCCCCCGTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1123

QY 1889 TCATATTTCCCTGTGAGGTTGCTCAGCAGCTGTGGCAAGGCA 1933

Db 1124 TCCAAATTTCCCTGGGAAGGTTGCTCACCACCTGTGGCAGGGGAA 1168

RESULT 3

BO707076

LOCUS

DEFINITION

AGENCOURT_8347306 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6279295

5', mRNA sequence.

ACCESSION

BO707076

VERSION

BO707076.1

KEYWORDS

EST.

SOURCE

human.

BO707076 884 bp mRNA linear EST 16-JUL-2002

AGENCOURT_8347306 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6279295

5', mRNA sequence.

ACCESSION BO707076

VERSION BO707076.1

KEYWORDS GI:21845975

EST.

SOURCE human.


```
Db 1 CACTTTTCTCGCTAGACGAGCAGCTGGAACAGACCCAGGCGGAGCACACCTGT 60
QY 187 GGGGGAGGAGCGCTCGAGAGCTTAGAGACCCAGCCGCGGTGATCTCACCATGTGC 246
Db 61 GGGGGAGGAGCGCTCGAGAGCTTAGAGACCCAGCCGCGGTGATCTCACCATGTGC 120
QY 247 GGATTTCCGAGCGCGCTCGAGAGCTTAGAGATCCGGAAGCACACCCGCGGTGTC 306
Db 121 GGATTTCCGAGCGCGCTCGAGAGCTTAGAGATCCGGAAGCACACCCGCGGTGTC 180
QY 307 GAAGCCACCAAGACTCGCGCTCTTGGAGAAAGCGTGAGCAGGGGCGCCAGCGGTCTCCG 366
Db 181 GAASCACCAA----- 190
QY 367 CGCCTGTCTGCACCCCTGCTGAGCTGCTGACAGTGACATGACATCCAGTTACCA 426
Db 191 -----ACATCCAGTTACCA 205
QY 427 GTGTCTTGAATTGATAGTGGCTTCTGTTGTGTCAGTCTCATATAAGAACTACAGCTCATC 486
Db 206 GTGTCTTGAATTGATAGTGGCTTCTGTTGTGTCAGTCTCATATAAGAACTACAGCTCATC 265
QY 487 AGGAGGAGATCGCAGCAGGCTAAGACACCAACACCATGTTTGCACGAGCTCAAGGA 546
Db 266 AGGAGGAGATCGCAGCAGGCTAAGACACCAACACCATGTTTGCACGAGCTCAAGGA 325
QY 547 TCTCAAGATCACAGGAGGTGCTCTTCTCTTACTGGCCAGGTCAGTCTCTCAACGA 606
Db 326 TCTCAAGATCACAGGAGGTGCTCTTCTCTTACTGGCCAGGTCAGTCTCTCAACGA 385
QY 607 GTCTTCAGAGGAGCAGCAGGAGTCTCTTCTCTTACTGGCCAGGTCAGTCTCTCAACGA 666
Db 386 GTCTTCAGAGGAGCAGCAGGAGTCTCTTCTCTTACTGGCCAGGTCAGTCTCTCAACGA 445
QY 667 AGCATTTCTGAGNAGAACATACAGAAAGTCTTCTCAAGAAACACCAAGTGGAGCCG 726
Db 446 AGCATTTCTGAGNAGAACATACAGAAAGTCTTCTCAAGAAACACCAAGTGGAGCCG 505
QY 727 AGTCTATCTTACACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCAGAGTTTGAACG 786
Db 506 AGTCTATCTTACACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCAGAGTTTGAACG 565
QY 787 GCTGAATGTGCACTTCAGAGAACATTTGGCAAGCACAATAAAGAAAGCAGGAAATC 846
Db 566 GCTGAATGTGCACTTCAGAGAACATTTGGCAAGCACAATAAAGAAAGCAGGAAATC 625
QY 847 TTTTGAAGAGAGACATTTGAAACAAATTTGCAGACCAAGCAGTTCACAGCAGGATTC 906
Db 626 TTTTGAAGAGAGASACTTTGAAACAAATTTGCAGACCAAGCAGTTCACAGCAGGATTC 685
QY 907 AGTGGAGGTTATCAAGAAATCTCTTGTGAGAGGTTTTTAAATAATGTTTACGAGGAAGA 966
Db 686 AGTGGAGGTTATCAAGAAATCTCTTGTGAGAGGTTTTTAAATAATGTTTACGAGGAAGA 745
QY 967 TGAACATCCTTGGGGTGTGGAGCACCCTTTAAAGATTTTTTAAACAGCTTCAGTAC 1026
Db 746 TGAACATCCTTGGGGTGTGGAGCACCCTTTAAAGATTTTTTAAACAGCTTCAGTAC 805
QY 1027 CTTTCTCAACAGAGACCATTTGCCAAGAGCAGGAAAGGGG 1071
Db 806 CTTTCTCAACAGAGACCATTTGCCAAGAGCAGGAAAGGGG 850
```

```
RESULT 5
LOCUS BG220276 777 bp mRNA linear EST 21-APR-2001
DEFINITION R640049 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG220276
VERSION BG220276.1 GI:13746297
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,
E., Veloso, N., Kika, A., Hess, J., Cothren, K., Lo, K., Offenbacher,
J., Danzig, J. and Ducar, M.
TITLE
Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL
MEDLINE
COMMENT
21227151
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 501.
FEATURES
source
1. .777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression'
Libraries using Random Activation of Gene Expression",
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 239 a 165 c 168 g 205 t
ORIGIN
Query Match 23.98; Score 720.8; DB 12; Length 777;
Best Local Similarity 97.58; Pred. No. 8e-131;
Matches 753; Conservative 0; Mismatches 17; Indels 2; Gaps 2;
QY 839 AGGAATCTTTGGAAGAGAGAGACTTTGAAAAACAAATTCAGAGCAGCAGTTGCAGCA 898
Db 4 AGGAATCTTTGGAAGAGAGAGACTTTGAAAAACAAATTCAGAGCAGCAGTTGCAGCA 63
QY 899 GGAGTTCAGTGGAGGTTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAAAATATGTTAC 958
Db 64 GGAGTTCAGTGGAGGTTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAAAATATGTTAC 123
QY 959 GAGGAGATCAAAACATCCTTGGGTGGTGGAGGACCCCTTAAAGATTTTTTAAACAGC 1018
Db 124 GAGGAGATCAAAACATCCTTGGGTGGTGGAGGACCCCTTAAAGATTTTTTAAACAGC 183
QY 1019 TTCAGTACCCTTCTGAAACAGAGCAGCATTGCCAAGAGCAGGAAAGGGCAGGCTT 1078
Db 184 TTCAGTACCCTTCTGAAACAGAGCAGCATTGCCAAGAGCAGGAAAGGGCAGGCTT 243
QY 1079 GAGGAGCCTCCATTCTATGCTGGATGAAGAGATGATTTCTACATGTTTACTACTTC 1138
Db 244 GAGGAGCCTCCATTCTATGCTGGATGAAGAGATGATTTCTACATGTTTACTACTTC 303
QY 1139 TTCCTTAAGAGAACACCTCCCTGATCTTCCCGGCATCATAAAGCAGCTGCTCAGTA 1198
Db 304 TTCCTTAAGAGAACACCTCCCTGATCTTCCCGGCATCATAAAGCAGCTGCTCAGTA 363
QY 1199 TTATATGAACGGAAGTGGAAAGTGTGTTAATGCCCTCCCTGCTTCCATATGATTGCGAGC 1258
Db 364 TTATATGAACGGAAGTGGAAAGTGTGTTAATGCCCTCCCTGCTTCCATATGATTGCGAGC 423
QY 1259 GAGTTGTGAATCAGCCCTTACTTGTGTACTCCGTTTCAATGAAGAAAGCAGCAGCAGTCC 1318
Db 424 GAGTTGTGAATCAGCCCTTACTTGTGTACTCCGTTTCAATGAAGAAAGCAGCAGCAGTCC 483
QY 1319 CTGTCCCGCCAGCAACCCAGTCTGCTGCTGATTTCCACATCGCTATTTCTTCAAGACA 1378
Db 484 CTGTCCCGCCAGCAACCCAGTCTGCTGCTGATTTCCACATCGCTATTTCTTCAAGACA 543
```

```
QY 1379 TTTCCATTCCATTTCATGTTTGACAAAGATATGACAATTCGCAATTTGGCAATGCGATC 1438
|||||
Db 544 TTTCCATTCCATTTCATGTTTGACAAAGATATGACAATTCGCAATTTGGCAATGCGATC 603
|||||
QY 1439 AGAAGGCTCATGAACAGGAGAGACTTTCAGGAAGAGCCCTAAATTTCAAGAAATACCTTGAA 1498
|||||
Db 604 AGAAGGCTCATGAACAGGAGAGACTTTCAGGAAGAGCCCTAAATTTCAAGAAATACCTTGAA 663
|||||
QY 1499 ATTCTGACTCCAAAATCAACAGAGAGCTTTAGCGGATCATGACTATGTTGAATATGCAG 1558
|||||
Db 664 ATTCTGACTCC-AAAATCAACAGAGAGCTTTAGC-GGAACATGACTATTTCTTAATATGCC 721
|||||
QY 1559 TTTGTTGACGAGTGAAGAGAGAGGACAACTCTGTGAAGAAATCTTCAAGG 1610
|||||
Db 722 GTTGTGTACTAGTAGGAAGATGGGACCACCTCTGTAAGAAATCTTCAAGG 773
|||||

RESULT 6
AL563000/c 829 bp mRNA linear EST 16-FEB-2001
LOCUS AL563000 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC027YN16 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563000
VERSION AL563000.1 GI:12911980
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 829)
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC027YN16"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com
http://fulllength.invitrogen.com"

BASE COUNT 224 a 170 c 165 g 267 t 3 others
ORIGIN
Query Match 23.5%; Score 707.4; DB 9; Length 829;
Best Local Similarity 99.3%; Pred. No. 3.3e-128;
Matches 708; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 905 CCAGTGGAGGTATCAAGAATCTCTGTGGTGAAGAGCTTTTAAAAATATGTTACGAGAA 964
|||||
Db 829 CCAGTGGAGGTATCAARAATCTCTGTGGTGAAGAGCTTTTAAAAATATGTTACGAGAA 770
|||||
QY 965 GATGAAGAAATCCTTCGGGTGGTGGAGGACCCCTTAAAGATTTTAAACAGCTTCACT 1024
|||||
Db 769 GATGAAGAAATCCTTCGGGTGGTGGAGGACCCCTTAAAGATTTTAAACAGCTTCACT 710
|||||
QY 1025 ACCCTTCTGAACAGAGACCCATTGCCAAGAGAGAGAAAGGGCGAGCTTGAGGAC 1084
|||||
```

```
Db 709 ACCCTTCTGAACAGAGACCCATTGCCAAGAGAGAGAAAGGGCGAGCTTGAGGAC 650
|||||
QY 1085 GCCTCCATTCTATGCTCGGATAGGAGGATGATTTTCTACATGTTTCTACTTCTTCCCT 1144
|||||
Db 649 GCCTCCATTCTATGCTCGGATAGGAGGATGATTTTCTACATGTTTCTACTTCTTCCCT 590
|||||
QY 1145 AAGAGAACACCTCCCTCGATTCTTCCCGGCATCATAAAGGCAGCTGCTCACGTATTATAT 1204
|||||
Db 589 AAGAGAACACCTCCCTCGATTCTTCCCGGCATCATAAAGGCAGCTGCTCACGTATTATAT 530
|||||
QY 1205 GAACGGAAGTGGAGTGCCTTAATGCTCCCTGCTTCCATATGATTCAGCGAGTTT 1264
|||||
Db 529 GAACGGAAGTGGAGTGCCTTAATGCTCCCTGCTTCCATATGATTCAGCGAGTTT 470
|||||
QY 1265 GTGAATCAGCCCTACTTGTGTACTCCGTTACATGAAAGCACCAGGCATCCCTGTCC 1324
|||||
Db 469 GTGAATCAGCCCTACTTGTGTACTCCGTTACATGAAAGCACCAGGCATCCCTGTCC 410
|||||
QY 1325 CCCAGAAACCCAGTCTCTCGTGTGATTCACATCGCTATTCGCAAGACATTTCCA 1384
|||||
Db 409 CCCAGAAACCCAGTCTCTCGTGTGATTCACATCGCTATTCGCAAGACATTTCCA 350
|||||
QY 1385 TTCCATTTTCATGTTTGACAAAGATATGACAATTCGCAATTTGGCAATGCGATCAGAAG 1444
|||||
Db 349 TTCCATTTTCATGTTTGACAAAGATATGACAATTCGCAATTTGGCAATGCGATCAGAAG 290
|||||
QY 1445 CTGATGAACAGAGAGACTTTCAGGAAAGCCCTAAATTTGAAGAACTTTGAAATCTG 1504
|||||
Db 289 CTGATGAACAGAGAGACTTTCAGGAAAGCCCTAAATTTGAAGAACTTTGAAATCTG 230
|||||
QY 1505 ACTCCAAAATCAACAGAGAGCTTTAGCGGATCATGACTATGTTGAATATGCAAGTTGTT 1564
|||||
Db 229 ACTCCAAAATCAACAGAGAGCTTTAGCGGATCATGACTATGTTGAATATGCAAGTTGTT 170
|||||
QY 1565 GTACGAGTGAAGAGATGGGACAACTCTGTGAAGAAATCTTCAAGGGTATGGA 1617
|||||
Db 169 GTACGAGTGAAGAGATGGGACAACTCTGTGAAGAAATCTTCAAGGGTATGGA 117
|||||

RESULT 7
AL570204/c 852 bp mRNA linear EST 16-FEB-2001
LOCUS AL570204 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI005YI13 3
DEFINITION prime, mRNA sequence.
ACCESSION AL570204
VERSION AL570204.1 GI:12926285
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 852)
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI005YI13"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
```


QY	2400	TCAAAGACATGCTCTGGTTTCGTGTTTACCCCTCGATCAAGGGAGGAACATCTCCACCAAACT	2459
Db	385	TCAAAGACATGCTCTGGTTTCGTGTTTACCCCTCGATCAAGGGAGGAACATCTCCACCAAACT	326
QY	2460	TCCTTAGTGAATCCCGCGAATCGCCATTTCTGGATGCTTACCACACAGGAACAACT	2519
Db	325	TCCTTAGTGAATCCCGCGAATCGCCATTTCTGGATGCTTACCACACAGGAACAACT	266
QY	2520	CAAAACCATGCTTCCAAAAGAAAGATGTGGAAGATGGCAATGGCAATTTTTTAGGCAAAAG	2579
Db	265	CAAAACCATGCTTCCAAAAGAAAGATGTGGAAGATGGCAATGGCAATTTTTTAGGCAAAAG	206
QY	2580	CATCAGAAATAGATTAGCAACCTATATACCTATTTAATAGTCTTTGGGTTTGACATCAATT	2639
Db	205	CATCAGAAATAGATTAGCAACCTATATACCTATTTAATAGTCTTTGGGTTTGACATCAATT	146
QY	2640	GAAGATGTGTAGAGCCTCTGAAAGCACTTTAGGGATTGTAGATGGCTAACACAGCATATT	2699
Db	145	GAAGATGTGTAGAGCCTCTGAAAGCACTTTAGGGATTGTAGATGGCTAACACAGCATATT	86
QY	2700	AAAATTTCAGGACCCAAAGTCACAACTTTCTCTGTTTAAACATGACAAA-----ATGTACT	2755
Db	85	AAAATTTCAGGACCCAAAGTCACAACTTTCTCTGTTTAAACATGACAAAATGTATGTACT	26
QY	2756	CACCTTCAGTACTTCAGCTCTTCAA	2779
Db	25	CACCTTCAGTACTTCAGCTCTTCAA	2

RESULT 9	AA55318/c	708 bp	linear	EST 09-SEP-1997
LOCUS	AA55318/c			
DEFINITION	nk79f06.s1 NCI_CGAP_Sch1 Homo sapiens cDNA clone IMAGE:1019747 3' similar to gb:X66334_cds1 similar to gb:X66334_cds1 (HUMAN);, mRNA sequence.			

ACCESSION	AA555318
VERSION	AA555318.1
KEYWORDS	GI:2325857
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ccgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: ccgaps-f@mail.nih.gov
COMMENT	Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1229 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 480.

```
FEATURES
source
1. .708
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:1019747
/clone_lib="NCI_CGAP_Sch1"
/tissue_type="Schwannoma tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK; Site_1: EcoRI; Site_2: XhoI
Cloned unidirectionally. Primer: Oligo dt. Two pooled
bulk Schwannoma tumors. 5' adaptor sequence: 5'
```

GAATTCGCACGAG	3'	3'	adaptor sequence: 5'
CTCGAGTTTTTTTTTTTT	3'	Average insert size:	1.2 kb.
198 a	153 c	150 g	207 t
BASE COUNT			
ORIGIN			

Query Match	Score 692.2; DB 9;	Length 708;
Best Local Similarity	99.4%;	
Pred. No. 3.2e-125;		
Mathews 705; Conservative		
0.1; Minotaur-b-32		

QY 2072 TGTGGAGCTGGATGTCTACAAGGTGGAGACCATTGGCGATGCCCTATTGTGTAGCTGGG 2131

Db 708 TGTGGAGCTGAATGCTCTACAAGTGGAGACCATTTGCGATGCCATTGTGTAGCTGGG 649

2132 GGATTACACAAAGAGAGTGATACTCATGCTGTTTCAGATAGCGCTGATGCCCTGAAGATG 2191

DD 048 GGATTTACACATAGAGAGTGTHACTCATGCTGTTGATGGCCCTGAAGATG 589

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

C

Db
528 CTGCACTCTGGATCAGTCTTTTCTGGCGTTCGTTGGAGTTAAATGCCCGCTACTGTCTT 469

QY 2312 TTTGGAAACAATGTCACTCTGGCTAACAAATTTGAGTCTGCAGTGTAACGAAAATC 2371

Db 468 TTGGAAACAATGTCACCTCTGGCTAACAAATTTGAGTCCTGCAGGTACCAACGAAAAATC 409

QY 2372 AATGTCAGCCCCAACAACTTACAGATTACTCAAAGACTGTCTGGTTTCGTGTTACCCCT 2431

408 CATGTCAGGCCCAACAACCTTACAGATTACTCAAAGACTGTCCTGGTGTTCGTGTGTACCCCT 349

[illegible][illegible]

288 CTGGATGCTTACCAACAAGGAACAACCTCAAAACCATGCTTCCAAAGCAACATCTGGAA 329

QY 2552 GATGGCAATGCCAATTTT TAGGCAAAAGCATCAGGAATAGATTAGCAACCTATATACCTA 2611

Db 228 GAT -GCAATGCCAATTTT TAGGCAAAGCATCAGGAATAGATTAGCAACCTATATACCTA 170

QY 2612 TTTATAAGTCTTTTGGGGTTTGACTCATTTGAAGATGTGTAGAGCCTCTGAAAGCACTTAG 2671

DB 169 TTTATAAGTCTTTGGGGTTTGACTCATTTGAAGATGTGTAGAGCCCTCTGAAAGCACTTTAG 110

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Db
49 CTGTTTAAACATGACAAAATGTACTCACTTCAGTACTTCAGTCTTCAAG 1

BQ774520/c

DEFINITION	UI-H-EZ1-bcb-b-11-0-UI.s1 NCI_CGAP_Ch2 Homo sapiens cDNA clone
-------------------	--

ACCESSION	BQ774520
VERSION	BQ774520.1
	GT.21982996

KEYWORDS
EST.
SOURCE
human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 741)

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of Orthopedics
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: M13 FORWARD
 Polya+yes.

FEATURES

Location/Qualifiers
 1. 741
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-H-E21-bcb-b-11-0-UI"
 /clone_lib="NCI_CGAP_Ch2"
 /tissue_type="Chondrosarcoma Grade II"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Organ: Left pelvis; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site: 1: EcoR I; Site 2: Not I; NCI_CGAP_Ch2 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma Grade II. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGATCAGCT.
 TAG_L1B-UI-H-E21
 TAG_TISSUE-grade-2-chondrosarcoma
 TAG_SEQ-AFCTAATG*
 BASE COUNT 198 a 157 c 156 g 227 t 3 others
 ORIGIN

Query Match 22.88; Score 688.2; DB 14; Length 741;
 Best Local Similarity 98.78; Pred. No. 1.9e-124;
 Matches 734; Conservative 0; Mismatches 6; Indels 4; Gaps 4;
 QY 2051 TACACTCGCTTCGACCAGCAGTGTGGAGAGCTGGATGCTTACAAAGTGGAGACCATTTGGC 2110
 DB 741 TACACTCGCTTCGACCAGCAGTGTGGAGAGCTGGATGCTTACAAAGTGGAGACCATTTGGC 684
 QY 2111 GATGCTTATTTGTAGTGTGGGGAT-TACACAAGAGAGTGTACTCATGCTCTTTCAGAT 2169
 DB 683 GATGCTTATTTGTAGTGTGGGGATACACNAAAAGAGAGTGTACTCATGCTCTTTCAGAT 624
 QY 2170 AGCGCTGATGGCCCTGAAGATGATGGAGCTCTCTGATGAATGATGCTCTCCCATGGAGA 2229
 DB 623 AGCGCTGATGGCCCTGAAGATGATGGAGCTCTCTGATGAATGATGCTCTCCCATGGAGA 565
 QY 2230 ACCTATCAAGATCGGAATTTGGATGCTGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 2289
 DB 564 ACCTATCAAGATCGGAATTTGGATGCTGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 505
 QY 2290 TAAATGCCCGTTACTGTCTTTTGGAAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2349
 DB 504 TAAATGCCCGTTACTGTCTTTTGGAAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 445
 QY 2350 CTCAGTGTACCAAGAAATCAATGTCAGCCCAACAACCTACAGATTACTCAAGACATC 2409
 DB 444 CTCAGTGTACCAAGAAATCAATGTCAGCCCAACAACCTACAGATTACTCAAGACATC 385
 QY 2410 TCCTGTTTCTGTTTACCCCTCGATCAAGGGAGGAACCTCCACCAACCTTCCTCCTAGTGA 2469
 DB 384 TCCTGTTTCTGTTTACCCCTCGATCAAGGGAGGAACCTCCACCAACCTTCCTCCTAGTGA 325

QY 2470 AATCCCGGAATCTGCCATTTCTGGATGCTTACCAACAAGGAACAACCTCAAAACCATG 2529
 DB 324 AATCCCGGAATCTGCCATTTCTGGATGCTTACCAACAAGGAACAACCTCAAAACCATG 265
 QY 2530 CTTCCAAAAGAAAGATGTGGAGATGCAATGCAATTTTGGGCAAGCATCAGGAAT 2589
 DB 264 CTTCCAAAAGAAAGATGTGGAGATGCAATGCAATTTTGGGCAAGCATCAGGAAT 205
 QY 2590 AGATTAGCAACCTATATACCTATTATTAAGCTCTTTGGGGTTTGGCTCATTTGAAGATGCT 2649
 DB 204 AGATTAGCAACCTATATACCTATTATTAAGCTCTTTGGGGTTTGGCTCATTTGAAGATGCT 145
 QY 2650 AGAGCCTCTGAAAGCAGCTTTAGGATGTAGATGGCTTAACAAGCAGTATTAAATTTTCAG 2709
 DB 144 AGAGCCTCTGAAAGCAGCTTTAGGATGTAGATGGCTTAACAAGCAGTATTAAATTTTCAG 85
 QY 2710 GAGCCAAGTCACAATCTTTCTCTCTGTTTAAACATGACAAAATGTACTCAGTACTTTC 2769
 DB 84 GAGCCAAGTCACAATCTTTCTCTCTGTTTAAACATGACAAAATGTACTCAGTACTTTC 25
 QY 2770 AGCTCTTCAAGAAAAA 2793
 DB 24 AGCTCTTCAAGAAAAA 1
 RESULT 11
 BC012627
 LOCUS BC012627 1288 bp mRNA linear HTC 20-AUG-2001
 DEFINITION Homo sapiens, Similar to guanylate cyclase 1, soluble, alpha 3,
 clone IMAGE:4295630, mRNA.
 ACCESSION BC012627
 VERSION BC012627.1 GI:15215000
 KEYWORDS HTC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1288)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (15-AUG-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Kettman and Anuradha Madan
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAL Plate: 19 Row: 0 Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 3702146
 This clone has the following problem: incomplete processing.
 FEATURES
 Location/Qualifiers
 1. 1288
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4295630"
 /tissue_type="Prostate"
 /clone_lib="NIH_MGC_83"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 BASE COUNT 408 a 278 c 291 g 311 t

ORIGIN

Query Match 22.8%; Score 687; DB 11; Length 1288;
Best Local Similarity 98.7%; Pred. No. 3e-124;
Matches 704; Conservative 0; Mismatches 5; Indels 4; Gaps 1;

QY 2094 AGGTGAGACCATGGCGATGCTATTGTGTAGCTGGGGATACACAAGAGAGTGATA 2153
DB 573 AGGTGAGACCATGGCGATGCTATTGTGTAGCTGGGGATACACAAGAGAGTGATA 632
QY 2154 CTCATCTCTGTTCAGATAGCGCTGATGGCCCTGAAGATGATGGAGCTCTCTGATGAAGTGA 2213
DB 633 CTCATCTCTGTTCAGATAGCGCTGATGGCCCTGAAGATGATGGAGCTCTCTGATGAAGTGA 692
QY 2214 TGTCTCCCATGAGAACCTATCAAGATGCGAATGAGTGCACCTCTGGATCATGTTTTTG 2273
DB 693 TGTCTCCCATGAGAACCTATCAAGATGCGAATGAGTGCACCTCTGGATCATGTTTTTG 752
QY 2274 CTGGCTCTGTGGAGTAAATGCCCGTACTGTCTTTTGGAAACAATGTCACCTCTGG 2333
DB 753 CTGGCTCTGTGGAGTAAATGCCCGTACTGTCTTTTGGAAACAATGTCACCTCTGG 812
QY 2334 CTAACAAATTTGAGTCTGCTGCTGATACACGAAATCAATGTCAGCCCAACAACCTTACA 2393
DB 813 CTAACAAATTTGAGTCTGCTGCTGATACACGAAATCAATGTCAGCCCAACAACCTTACA 872
QY 2394 GATTACTCAAGACTCTCTGTTTGTGTTTACCCCTCGATCAAGGGAGGAACTTCCAC 2453
DB 873 GATTACTCAAGACTCTCTGTTTGTGTTTACCCCTCGATCAAGGGAGGAACTTCCAC 932
QY 2454 CAAACTCCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGTTTACCACAGGAA 2513
DB 933 CAAACTCCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGTTTACCACAGGAA 992
QY 2514 CAAACTCAAAACCATGCTTCCAAAGAAAGATGTGAAGATGCAATGCCAATTTTTTAG 2573
DB 993 CAAACTCAAAACCATGCTTCCAAAGAAAGATGTGAAGATGCAATGCCAATTTTTTAG 1052
QY 2574 GCAAAGATCAGGAATAGATTAGCAACCTATATACCTATTATTAAGTCTTTGGGGTTGA 2633
DB 1053 GCAAAGATCAGGAATAGATTAGCAACCTATATACCTATTATTAAGTCTTTGGGGTTGA 1112
QY 2634 CTCATTGAAGATGTAGAGCTCTGAAGCACTTTAGGATTTAGATGGCTTAAACAGC 2693
DB 1113 CTCATTGAAGATGTAGAGCTCTGAAGCACTTTAGGATTTAGATGGCTTAAACAGC 1172
QY 2694 AGTATTAAATTTTCAGGAGCAAGTCAATCTTCTCTGTTTAAACATGACAAA---A 2749
DB 1173 AGTATTAAATTTTCAGGAGCAAGTCAATCTTCTCTGTTTAAACATGACAAAATGTA 1232
QY 2750 TGTACTCAGTCTAGTACTTCTAGCTCTTCAAGAAAAAATAACCTTAAAAA 2802
DB 1233 TGTACTCAGTCTAGTACTTCTAGCTCTTCAAGAAAAAATAACCTTAAAAA 1285

RESULT 12

BI771039
LOCUS 60305388F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204997 5',
DEFINITION mRNA sequence.

ACCESSION BI771039.1 GI:15762617

VERSION 1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 824)

NIH-MGC <http://mgs.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

FEATURES

Source
1. 824
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5204997"
/lab_host="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT 250 a 189 c 216 g 169 t
ORIGIN

Location/Qualifiers

1. 824
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5204997"
/lab_host="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT 250 a 189 c 216 g 169 t
ORIGIN

Query Match. 22.3%; Score 673.4; DB 13; Length 824;
Best Local Similarity 88.8%; Pred. No. 1.5e-121;
Matches 814; Conservative 0; Mismatches 6; Indels 97; Gaps 4;

QY 176 AGGACACCTGTGGGGAGGAGCGCTGGAGGAGCTTAGAGACCCAGCGCGGTGATC 235
DB 1 AGGACACCTGTGGGGAGGAGCGCTGGAGGAGCTTAGAGACCCAGCGCGGTGATC 60
QY 236 TCACCATGTGCGGATTTGGAGGCGCCCTTGAGTGTCTAGAGTCCGGAAGCACAGCC 295
DB 61 TCACCATGTGCGGATTTGGAGGCGCCCTTGAGTGTCTAGAGTCCGGAAGCACAGCC 120
QY 296 CCGAGTGTGCGAAGCACCACCAAGACTGCGGCTCTTGGAGAAAGCGTGAGCGGGCCAC 355
DB 121 CCGAGTGTGCGAAGCACCACCAAGACTGCGGCTCTTGGAGAAAGCGTGAGCGGGCCAC 142
QY 356 CCGGCTCTCGCGCCTGTCTGACCCCTGTGCGCTGAGCTGCCTGACAGTGACAATGACAT 415
DB 143 -----ACAT 146
QY 416 CCAGATTACAGTGTCTTGAATTGATAGTGGTCTCTGTTGTGAGTCTCATATAAGAAC 475
DB 147 CCAGATTACAGTGTCTTGAATTGATAGTGGTCTCTGTTGTGAGTCTCATATAAGAAC 206
QY 476 TACAGTCTCAGGAGGAGATCGCAGCAGGTTAAGAGACACCAACCATGTTCTGCAGC 535
DB 207 TACAGTCTCAGGAGGAGATCGCAGCAGGTTAAGAGACACCAACCATGTTCTGCAGC 266
QY 536 AAGCTCAAGGATCTCAAGATCAGAGGAGTGTCTTCTCTTACTGGCACCAGGTCAC 595
DB 267 AAGCTCAAGGATCTCAAGATCAGAGGAGTGTCTTCTCTTACTGGCACCAGGTCAC 326
QY 596 GTTCTTAAGAGTCTTTCAGAGGAGGAGCAGGAGGTCAGAGAGTCTGCAAGCAACCGTG 655
DB 327 GTTCTTAAGAGTCTTTCAGAGGAGGAGCAGGAGGTCAGAGAGTCTGCAAGCAACCGTG 386
QY 656 CCCATCTGCAACACATCTCTGAGAGAACATACAGAAAGTCTTCTCTCAAGAAACCC 715
DB 387 CCCATCTGCAACACATCTCTGAGAGAACATACAGAAAGTCTTCTCTCAAGAAACCC 446
QY 716 AGTCGGAGCGGAGTCTATCTTACACTTTTGGCAGAGAGTATTGGCAACTGATTTCCCA 775
DB 447 AGTCGGAGCGGAGTCTATCTTACACTTTTGGCAGAGAGTATTGGCAACTGATTTCCCA 506

776	QY	GAGTTGAACGGCTGAATGTTGCACATTCAGAGAACAATTCGCAAGCACAAAATAAAGAA	835
507	Db	GAGTTTGAACGGCTGAATGTTGCACATTCAGAGAACAATTCGCAAGCACAAAATAAAGAA	566
836	QY	AGCA - CGAATATCTTTGGAAGAGAAGACATTTGAAAAACAATTCGACAGCAAGCAGTTGC	894
567	Db	AGCACGGAATCTTTGGAAAGAGAAGACTTTGAAAAACAATTCGACAGCAAGCAGTTGC	626
895	QY	ACGACGAGTTCACAGTCGAGGTTATCAAGAATCTCTTGGTGAAGAGGTTTTTAAATATG	954
627	Db	AGCAGGAGTTCACAGTCGAGGTTATCAAGAATCTCTTGGTGAAGAGGTTTTCTAAATATG	686
955	QY	TTACGAGGAGAGATGAAAAACATCTCTGGGTGGTTGGAGGCACCCCTTAAGCATTTTTTAAA	1014
687	Db	TTACGAGGAGAGATGAAAAACATCTCTGGGTGGTTGGAGGCACCCCTTAAGCATTTTTTAA	746
1015	QY	CAG - CTTTCAGTACCCCTTCTGAAACAGACAGCACGCCATTTGCCAAGAGCAGGAAAAAGGGCCA	1073
747	Db	CAGCCCTTCAGTACCCCTTCGAAACAGACAGCAGCCATTTGCCAAGAGCAGGAAACACGGGGCA	806
1074	QY	-GGCTTGAGGACGCCTC	1089
807	Db	CGCCTTTGAGGACGCCTC	823

BASE COUNT	242 a	182 c	171 g	211 t	
ORIGIN					
Query Match		22.2%	Score 670;	DB 12;	Length 806;
Best Local Similarity		98.1%	Pred. No. 6,7e-121;		
Matches 720;	Conservative	0;	Mismatches 10;	Indels 4;	Gaps 4;
QY	895	AGCAGGAGTTCAGTGGAGGTTATCAAGAATCTCTTGGTGAAGAGGTTTTAAAAATATG	954		
Db	1	AGCAGGAGTTCAGTGGAGGTTATCAAGAATCTCTTGGTGAAGAGGTTTTAAAAATATG	60		

RESULT 14	ACCESSION	REFERENCE	FEATURES
AL527890	VERSION	AUTHORS	source
LOCUS	KEYWORDS	TITLE	
DEFINITION	SOURCE	JOURNAL	
	ORGANISM	COMMENT	

955	QY	TTACGAGGAGATGAAAAATCCTTTGGGGTGTGGAGCACCCCTTAAGATTTTTTAAA	1014
61	Db	TTACGAGGAGATGAAAAATCCTTTGGGGTGTGGAGCACCCCTTAAGATTTTTTAAA	120
1015	QY	CAGCTTCAGTACCCTTCTCAAAACAGACGAGCCATTGCCAAGAGCAGAGAAAGGGCGAG	1074
121	Db	CAGCTTCAGTACCCTTCTGAAAACAGACGAGCCATTGCCAAGAGCAGAGAAAGGGCGAG	180
1075	QY	GCTTGAGGAGCGCTCCATCTTATGCTGTGATAAGGAGGATGATTTCTACATGTTTACTA	1134
181	Db	GCTTGAGGAGCGCTCCATCTTATGCTGTGATRAGGAGGATGATTTCTACATGTTTACTA	240
1135	QY	CTTCTTCCCTAAGAGAACACCTCCCTGATCTTCCCGGCATCATAAAGGCGAGCTGCTCA	1194
241	Db	CTTCTTCCCTAAGAGAACACCTCCCTGATCTTCCCGGCATCATAAAGGCGAGCTGCTCA	300
1195	QY	CGTATTATATGAACGGAAAGTGCAGTGTCTTAATGCCTCCCTGCTCCCATATGATTG	1254
301	Db	CGTATTATATGAACGGAAAGTGCAGTGTCTTAATGCCTCCCTGCTCCCATATGATTG	360
1255	QY	CAGCGAGTTTGTGAATACAGCCCTACTTGTGTGTACTCCGTTACATGAAAGCACCAGCC	1314
361	Db	CAGCGAGTTTGTGAATACAGCCCTACTTGTGTGTACTCCGTTACATGAAAGCACCAGCC	420
1315	QY	ATCCCTGTCCCCCAGCAAAACCCAGTCCCTGCTGTGTATGCCACATCGCTATTCTCGAA	1374
421	Db	ATCCCTGTCCCCCAGCAAAACCCAGTCCCTGCTGTGTATGCCACATCGCTATTCTCGAA	480
1375	QY	GACATTTCCATTTCCATTTTCATGTTTGACAAAGATATGACAAATTCGCAATTTGGCAATGG	1434
481	Db	GACATTTCCATTTCCATTTTCATGTTTGACAAAGATATGACAAATTCGCAATTTGGCAATGG	540
1435	QY	CATCAGAAGGCTGATGAACAGAGAGACTTTCACGAGAAAGCCTAATTTTGAAGATACTTT	1494
541	Db	CATCAGAAGGCTGATGAACAGAGAGACTTTCACGAGAAAGCCTAATTTTGAAGATACTTT	600
1495	QY	T-GAAATTCGACTCCAAAAATCAACAGACGCTTTAGCGGGATCATGACTATGTTGA-AT	1552
601	Db	TGCAAAATTCGACTCCAAAAATCAACAGACGCTTTAGCGGGATCATGACTATGTTGACAT	660
1553	QY	ATGCACTTTGTGTACGAGTGAAGGAG-ATGGGACAACTCTGTGA-GAATCTTCAAGGG	1610
661	Db	ATGCACTTTGTGTACGAGTGAAGGAGCTGGGACAACCTGTGTGAACCAAAATCTTCAAGGG	720
1611	QY	TTATGGACCTCAAA	1624
721	Db	GTACGGAAAAACATA	734

/db_xref="taxon:9606"
/clone="CS0DC027YN16"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"
BASE COUNT 231 a 173 c 203 g 159 t 1 others
ORIGIN
Query Match 21.7%; Score 655.2; DB 9; Length 767;
Best Local Similarity 88.4%; Pred. No. 5.3e-118;
Matches 762; Conservative 1; Mismatches 4; Indels 95; Gaps 1;
QY 114 CAGAGTTTCTTACACTTTCTCGCTAGACGAGCAGCCTGGAACACCCAGGC 173
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 CAGAGTTTCTTACACTTTCTCGCTAGACGAGCAGCCTGGAACACCCAGGC 60
QY 174 GGAGGACACTGTGGGGAGGAGGCGCTGGAGGAGCTTAGAGACCCAGCGGGCGTA 233
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 GGAGGACACTGTGGGGAGGAGGCGCTGGAGGAGCTTAGAGACCCAGCGGGCGTA 120
QY 234 TCTCACCATTGCGGATTTGCGAGGCGCCCTGGAGCTGTAGATCCGGAAGCACAG 293
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 TCTCACCATTGCGGATTTGCGAGGCGCCCTGGAGCTGTAGATCCGGAAGCACAG 180
QY 294 CCCCAGGCTGCGAAGCCACCAAGACTGCGGCTCTTGGAGAAGCGTGAGCGGGGCC 353
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 CCCCAGGCTGCGAAGCCACCA-----203
QY 354 ACCGCGTCTCCGCGCTGTCTGCACCCCTGTCGCTGAGCTGCTGACATGAC 413
Db ||||-----AC 205
QY 414 ATCCAGTTACCAGTGTCTTGAATTGATAGTGGCTTCTGTGTGCTCATATAAGA 473
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
206 ATCCAGTTACCAGTGTCTTGAATTGATAGTGGCTTCTGTGTGCTCATATAAGA 265
QY 474 ACTACAGTCTATCAGGAGGATCGCAGCAGGTAAGACACCAACACCATGTTCTGCA 533
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
266 ACTACAGTCTATCAGGAGGATCGCAGCAGGTAAGACACCAACACCATGTTCTGCA 325
QY 534 CGAAGTCAAGATCTCAAGATCACAGGAGAGTGCTTCTTCTTCTTCTTCTTCTTCT 593
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
326 CGAAGTCAAGATCTCAAGATCACAGGAGAGTGCTTCTTCTTCTTCTTCTTCTTCT 385
QY 594 AAGTTCTTAACAGTCTTACAGGAGGCGCAGCAGGAGCTCAGAGACTGCAAGCAACCG 653
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
386 AAGTTCTTAACAGTCTTACAGGAGGCGCAGCAGGAGCTCAGAGACTGCAAGCAACCG 445
QY 654 TGCCATCTGTCAAGACATCTCTGAGAACAACATCAAGAAAGTCTTCTCAAGAAAAA 713
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
446 TGCCATCTGTCAAGACATCTCTGAGAACAACATCAAGAAAGTCTTCTCAAGAAAAA 505
QY 714 CCAGTGGAGCGGAGTCTATCTTCACTTTTGGCAGAGAGTATTTTGCAAACTGATTTCC 773
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
506 CCAGTGGAGCGGAGTCTATCTACACTTTTGGCAGAGAGTATTTTGCAAACTGATTTCC 565
QY 774 CAGAGTTTGAAGCGCTGAATGTGCACTTCAGAGAAACATTTGCAAGACAAATAAAG 833
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
566 CAGAGTTTGAAGCGCTGAATGTGCACTTCAGAGAAACATTTGCAAGACAAATAAAG 625
QY 834 AAGCAGGAGAAATCTTTGGAAGAGAGACTTTCAAAAACAATTCAGAGCAAGCAGTGG 893
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 626 AAAGCAGGAAATCTTTGGAAAGAGAGACTTTGAAAAACAATTCAGAGCAGCAGTTG 685
QY 894 CAGCAGGAGTCCAGTGGAGGTTATCAAGAATCTCTTGTGAAGAGGTTTTTAAATAT 953
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
686 CAGCAGGAGTCCAGTGGAGGTTATCAAGAATCTCTTGTGAGAGGTTTTTAAATAT 745
QY 954 GTTACGAGGAAGATGAACAT 975
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
746 GTTACGAGGAGATGAACAT 767
RESULT 15
BQ549434/C
LOCUS
DEFINITION
BQ549434 658 bp mRNA linear EST 17-JUN-2002
1106h09.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029417
3' similar to SW:CYG3_HUMAN Q02108 GUANYLATE CYCLASE SOLUBLE,
ALPHA-1 CHAIN ; , mRNA sequence.
ACCESSION BQ549434
VERSION BQ549434.1 GI:21433939
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 658)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,
M., Gibbons,M., McCann,R., Cole,R., Tsagarelisvilli,R., Williams,T.,
Jackson,X. and Bowers,Y.
Endocrine Pancreas Consortium
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Unpublished (2000)
TITLE
JOURNAL
COMMENT
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center for information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 485.
FEATURES
Location/Qualifiers
1..658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6029417"
/clone_lib="Human insulinoma"
/tissue_type="Insulinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
XhoI; Site_2: EcoRI; Constructed with lambda ZAPRI system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
BASE COUNT 181 a 131 c 140 g 206 t
ORIGIN
Query Match 21.7%; Score 654.8; DB 14; Length 658;
Best Local Similarity 99.7%; Pred. No. 6.5e-118;
Matches 656; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2135 TTACACAAAGAGAGTGATCTATCTGTTTCAGATAGCGGTGATGGCCCTGAAGATGATG 2194
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
658 TTACACAAAGAGAGTGATCTATCTGTTTCAGATAGCGGTGATGGCCCTGAAGATGATG 599
Qy 2195 GAGCTCTCTGATGAAGTTATCTCTCCCATGGGAGAACCTATCAGATGCGAATGGACTG 2254
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
598 GAGCTCTCTGATGAAGTTATCTCTCCCATGGGAGAACCTATCAGATGCGAATGGACTG 539
Qy 2255 CACTCTGGATCAGTTTTTGGCTGGCTGTTGGAGTTAAATGCCCGCTTACTGTCTTTTTT 2314
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
538 CACTCTGGATCAGTTTTTGGCTGGCTGTTGGAGTTAAATGCCCGCTTACTGTCTTTTT 479
Qy 2315 GGAACAAATGTCACTCTGGCTTAACAAATTTGAGTCTGACAGTGTACACGAAATCAAT 2374
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
478 GGAACAAATGTCACTCTGGCTTAACAAATTTGAGTCTGACAGTGTACACGAAATCAAT 419
Qy 2375 GTCAGCCCAACAACTTACAGATTACTCAAGACTGTCTGGTTTCGTTTACCCCTCGA 2434
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
418 GTCAGCCCAACAACTTACAGATTACTCAAGACTGTCTGGTTTCGTTTACCCCTCGA 359
Qy 2435 TCAAGGGAGGAACCTCCACCAAACTTCCCTAGTGAATCCCGGAATCTGCCATTTTCTG 2494
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
358 TCAAGGGAGGAACCTCCACCAAACTTCCCTAGTGAATCCCGGAATCTGCCATTTTCTG 299
Qy 2495 GATGCTTACCAACAAAGGAACAACTCAAAACCATGCTTCCAAAAGAAAGATGTGGAAGAT 2554
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
298 GATGCTTACCAACAAAGGAACAACTCAAAACCATGCTTCCAAAAGAAAGATGTGGAAGAT 239
Qy 2555 GGCAATGCCAATTTTTTAGGCAAGCATCAGGATAGATTAGCAACCTATATACCTATTT 2614
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
238 GGCAATGCCAATTTTTTAGGCAAGCATCAGGATAGATTAGCAACCTATATACCTATTT 179
Qy 2615 ATAGTCTTTGGGCTTTGACTCAATTTGAAGATGTGTAGAGCTCTGAAAGCACCTTTAGGGA 2674
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
178 ATAGTCTTTGGGCTTTGACTCAATTTGAAGATGTGTAGAGCTCTGAAAGCACCTTTAGGGA 119
Qy 2675 TTGTAGATGGCTAACAAAGCAGTATTAATAATTCAGGAGCCAAAGTCACAATCTTTCTCCTG 2734
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
118 TTGTAGATGGCTAACAAAGCAGTATTAATAATTCAGGAGCCAAAGTCACAATCTTTCTCCTG 59
Qy 2735 TTTAACATGACAAATCTACTCTCAGTACTCTCAGCTCTTCAAGAAATAAAAAA 2792
Db ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
58 TTTAGCATGACAAATCTACTCTCAGTACTCTCAGCTCTTCAAGAAATAAAAAA 1

Search completed: July 1, 2003, 17:26:01
Job time : 2768.9 secs

FEATURE:
NAME/KEY: 5'clip
LOCATION: 1..10
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 3236...3784
US-07-623-033-1

Query Match 4.28; Score 125.6; DB 1; Length 3784;
Best Local Similarity 53.48; Pred. No. 6.4e-24;
Matches 313; Conservative 0; Mismatches 264; Indels 9; Gaps 2;
QY 1812 GGAAGCTGAAGGTTACCTTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAAAGA 1871
DB 2334 GGAACCTGGAGCACCTGGTGGAGGAAAGGACTCAGCTGTACAAAGCCGAGGACAGGG 2393
QY 1872 CAGTAGACCTCTGCTCCATATATTCCTGTGAGGTTGCTCAGCAGCTGGCAAGGGC 1931
DB 2394 CTGACCACTTAATCTTATGCTGCTCCACGGCTGGTGTAAAGTCCCTGAAGGAGAAAG 2453
QY 1932 AAGTTGTGCAAGCAAGGTTAGTAATGTACCATGCTCTTCTCAGACATCGTTGGGT 1991
DB 2454 GCATCTGGAGCAGAGCTGTACGAAGAAGTCAATCTATTTTCAGTGACATGTGCGTT 2513
QY 1992 TCAGTGCATCTGCTCCAGTGTCAACGCTGAGGTCATCACCATGCTCAATGCACTGT 2051
DB 2514 TCAGGACCATCTGCAAGTACAGCAGCCGCTGAGGTTGGTGGACATGCTGAATGACATCT 2573
QY 2052 ACAGTCCCTGCGACCAAGTGTGGAGAGCTGGAGTGTCTACAGGTTGGAGACCATGGCG 2111
DB 2574 ACAAGATTTTGACCAAGTGTGGATCACCAGCAGCTCTACAAGGTAGAAACCATGGCG 2633
QY 2112 ATGCTATTGTGAGCTGGGGGATTAC---ACAAAGAGAGTGATCTCATCTGCTTTCAGA 2168
DB 2634 ATGCTAGCTGTGGCCAGCGGCTGCTATGAGAACGCGCAACCGGCATGCGTGGACA 2693
QY 2169 TAGCGTGTAGGCGCTGAAGATGATGGAGTC-----TCTGATGAAGTTATGTCTCCC 2222
DB 2694 TTTTCAAGATGGGCTTGGATCTCAGCTTCATGGGGACCTTTGAGCTGGAGCATCTCC 2753
QY 2223 ATGAGAACATATCAAGATCGAATGCACTGCACTCTGATCATGATGTTTGTGGCGTCG 2282
DB 2754 CCGGCTCCCGTGTGATTCGATGAGGTTGATCTGCGCCCTGCTGCTGCTGGTGG 2813
QY 2283 TTGGATTAATAATGCCCGCTTACTGCTTTTGGAAACAATGTCACCTCTGGCTAACAAT 2342
DB 2814 TGGGATCAAGATGCCCTGTTATGCTGTTTGGAGACACTGTCAACACTGCTCCAGGA 2873
QY 2343 TTGAGTCTGCACTGTACCAAGAAATCAATGTACGCCCAACAC 2388
DB 2874 TGGAGTCCACCGGCTTCCCTTAAGGATTCACATGAGCAGCTCCAC 2919

RESULT 2
US-08-908-643C-85
Sequence 85, Application US/08908643C
Patent No. 6120995
GENERAL INFORMATION:
APPLICANT: Waldman, Scott A.
Pearlman, Joshua M.
Barber, Michael T.
Parkinson, Stephanie
Kinslow, Scott J.
TITLE OF INVENTION: COMPOSITIONS THAT SPECIFICALLY BIND TO
COLORECTAL CANCER CELLS AND METHODS OF
USING THE SAME
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6120995ris LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.

ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch disk, 1.44 Mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/908.643C
FILING DATE: 07-Aug-1997
CLASSIFICATION: N/A
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mark Deluca
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-2209
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 3603
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-08-908-643C-85

Query Match 4.08; Score 121.4; DB 3; Length 3603;
Best Local Similarity 53.28; Pred. No. 8.6e-23;
Matches 307; Conservative 0; Mismatches 261; Indels 9; Gaps 2;
QY 1812 GGAAGCTGAAGGTTACCTTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAAAGA 1871
DB 2224 GAAACCTGGAAACATCTGGTAGAGGAAAGGACACAGCTGTACAAGGAGAGGAGGACAGG 2283
QY 1872 CAGTAGACCTTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCTGTGGCAAGGGC 1931
DB 2284 CTCACAGACTTAATCTTATGCTTCCCAAGGCTAGTGGTAAAGTCTCTGAAGGAGAAAG 2343
QY 1932 AAGTTGTGCAAGCAAGGTTAGTAATGTACCATGCTCTTCTCAGACATCGTTGGGT 1991
DB 2344 GCTTTGGAGCGGAACTATATGAGGAAGTTTACAATCTTACTTCACTGACATGTAGGTT 2403
QY 1992 TCAGTGCATCTGCTCCAGTGTCAACGCTGAGGTCATCACCATGCTCAATGCACTGT 2051
DB 2404 TCAGTACTATCTCAATATCAGCAGCCGCTGAGGAGTGGTGGACATGCTTAATGACATCT 2463
QY 2052 ACAGTCCCTGCGACCAAGTGTGGAGAGCTGGATGTCTACAAGGTTGGAGACCATGGCG 2111
DB 2464 ATAAGAGTTTGGACCACTTGTGATCATCATGATGTCTACAAGTGGAAACCATCGGTG 2523
QY 2112 ATGCTATTGTGAGCTGGGGATTAC---ACAAAGAGAGTGATCTCATCTGCTTCCCG 2168
DB 2524 ATGCGTACATGTTGGCTAGTGTGCTTTCCTTAAGAGAAATGGCAATCGCATGCAATAGACA 2583
QY 2169 TAGCGTGTAGGCGCTGAAGATCATGAGCTC-----TCTGATGAAGTTATGCTTCCCG 2222
DB 2584 TTGCCAAGATGGCTTGGAAATCTCAGCTTCATGGGGACCTTTGAGCTGGAGCATCTTC 2643
QY 2223 ATGAGGAACCTATCAAGATGCGAATGGAGTGGACTGTCACCTCTGGATCAGTTTTTGTGGGCTG 2282
DB 2644 CTGGCTCCCAATATGATGATGCGATTTGGATTTGACCTCTGCTGCTGCTGCTGAGTTG 2703
QY 2283 TTGAGTTAAATGCCCGCTTACTGCTTTTGGAAACAATGTCACCTCTGGCTTAAACAAT 2342
DB 2704 TGGGAATCAAGATGCTGCTGTTATGCTTATTTGGAGATGAGTGTCAACACAGGCTCTAGGA 2763
QY 2343 TTGAGTCTGCACTGTACCAAGAAATCAATGTCTAG 2379
DB 2764 TGGAAATCCACTGGCTTCCCTTTGAGATTCACGTGAG 2800

RESULT 3

US-08-908-643C-83
; Sequence 83, Application US/08908643C
; Patent No. 6120995

GENERAL INFORMATION:

APPLICANT: Waldman, Scott A.
Pearlman, Joshua M.
Barber, Michael T.
Schultz, Stephanie
Parkinson, Scott J.

TITLE OF INVENTION: COMPOSITIONS THAT SPECIFICALLY BIND TO
COLORECTAL CANCER CELLS AND METHODS OF
USING THE SAME

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6120995rls LLP
STREET: One Liberty Place - 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch disk, 1.44 Mb

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/908,643C

FILING DATE: 07-Aug-1997

CLASSIFICATION: N/A

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mark Deluca

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-2209

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 83:

SEQUENCE CHARACTERISTICS:

LENGTH: 3645

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 83:

US-08-908-643C-83

Query Match 4.0%; Score 121.4; DB 3; Length 3645;

Best Local Similarity 53.2%; Pred. No. 8.6e-23;

Matches 307; Conservative 0; Mismatches 261; Indels 9; Gaps 2;

QY 1812 GGAAGCTGAAGCTACCTTGAGCAAGCCCAAGCCCTGAGGAGGAGAGAAAAGA 1871
DB 2293 GAAAGCTGACATCTGGTAGAGGAAGACACAGCTGTACAAAGCAGAGGGACAGG 2352
QY 1872 CAGTAGACCTTGTGCTCCATATTCCCTGTGAGGTTGCTCAGCAGCTGTGGCAAGGCG 1931
DB 2353 CTGACAGACTTAACCTTTATGTGCTTCCAAAGCTAGTGGTAAAGTCTCTGAAGGAGAAG 2412
QY 1932 AGTTGTGAGCCCAAGAGTTCAAGTATGTCACCATGCTCTTCACAGATCGTTGGGT 1991
DB 2413 GCTTTGTGAGCGCGGAACATATATGAGGAAGTTACAATCTACTTCAGTGACATTTAGGTT 2472
QY 1992 TCACGTGACATCTGCTCCAGTCTCAGGCTCAGGTCATCACCATGCTCATGTCACGT 2051
DB 2473 TCACGTGACATCTGCAATACAGCAGCCCATGGAAGTGGTGACATGCTTATGACATCT 2532
QY 2052 ACACGTGCTTCACCAAGCAGTGTGGAGAGCTGGATGCTTACAAAGTGGAGACCATTTGGG 2111
DB 2533 ATAAGAGTTTGACCAATGTTGATCATCATGATGCTTACAAAGTGGAAACCATCGGTG 2592

QY 2112 ATGCCTATTGTAGCTGGGGATTAC---ACAAAGAGAGTATGATCTATGCTGTTCAGA 2168
DB 2593 ATGCCTACATGTGTGCTAGTGTTCCTTAAGAGAAATGGCAATCGGCATGCAATAGACA 2652
QY 2169 TAGCCCTGATGCCCTGAAGATGATGGAGCTC-----TCTGATGAAGTTATGCTCTCCC 2222
DB 2653 TTGCAAGATGCCCTTGGAAATCCCTCAGCTTCAGGGGACCTTTGAGCTGGAGCATCTTC 2712
QY 2223 ATGGAAGACCTATCAAGATGCGAATGGACTGCACTCTGATCAGTTTTTGTGCTGGCTGG 2282
DB 2713 CTGGCTCCCAATATGGAATTCGCATTTGAGTTCACTCTGCTCCCTGCTGCTGAGTTG 2772
QY 2283 TTGGAGTTAAATGCCCGTTTACTGTCTTTTGGAAACAATGCTCACTCTGGCTAAACAAT 2342
DB 2773 TGGGAATCAAGATGCTGCTGTTATTGCTATTGAGATACGCTCAACACAGCCTCTAGCA 2832
QY 2343 TTGACTCTGCTGAGTGTACCAAGCAAAAATCAATGTCAG 2379
DB 2833 TGGAAATCCACTGGGCTCCCTTTGAGAATTCACGTGAG 2869

RESULT 4

US-08-908-643C-84

; Sequence 84, Application US/08908643C

; Patent No. 6120995

GENERAL INFORMATION:

APPLICANT: Waldman, Scott A.

Pearlman, Joshua M.

Barber, Michael T.

Schultz, Stephanie

Parkinson, Scott J.

TITLE OF INVENTION: COMPOSITIONS THAT SPECIFICALLY BIND TO

COLORECTAL CANCER CELLS AND METHODS OF

USING THE SAME

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6120995rls LLP

STREET: One Liberty Place - 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch disk, 1.44 Mb

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/908,643C

FILING DATE: 07-Aug-1997

CLASSIFICATION: N/A

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mark Deluca

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-2209

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:

LENGTH: 3745

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 84:

US-08-908-643C-84

Query Match

Best Local Similarity

4.0%; Score 121.4; DB 3; Length 3745;

53.2%; Pred. No. 8.7e-23;

Matches 307; Conservative 0; Mismatches 261; Indels 9; Gaps 2;

QY 1812 GGAAGCTGAAGGCTACCTTGGAGCAAGCCACCAAGCCCTGGAGGAGGAGAGAGAAAGA 1871
DB 2366 GAAACCTGGACATCTGGTAGAGAAAGGACACAGCTGTACAGGCGAGAGGACAGGG 2425
QY 1872 CAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTGTCTCAGACGCTGTGCAAGGGC 1931
DB 2426 CTGACACATTAACCTTATGCTTCCAGGCTAGTGGTAAAGTCTCTCAAGGAGAAAG 2485
QY 1932 AAGTTGTGCAAGCAAGAACTTCAATATGTCACCATGCTTCTCAGACATCTGTTGGGT 1991
DB 2486 GCTTTGGAGCGGAACTATATGAGAGTTTCAATCTACTCTCAGTGACATTTAGGTT 2545
QY 1992 TCAGTGCACTGTCTCCAGTGTCTACCGCTGAGGTCATCACATGCTCAATGCACTGT 2051
DB 2546 TCAGTACTATCTGCAATATACAGCACCCCATGGAAGTGGAGCATGCTTAAAGCATCT 2605
QY 2052 ACAGCTGCTCGACAGCAGTGTGGAGAGCTGTCTACAGGTGGAGACCATTTGGCG 2111
DB 2606 ATAAGAGCTTTTGACCACTATTTGATCATCATGATGTCTACAAGGTGGAACCATCGGTG 2665
QY 2112 ATGCTATTGTGAGCGGGGATTAC---ACAAAGAGATGATCATCTGCTGTTTACA 2168
DB 2666 ATGCTATACATGTTGGCTAGTGGTTGCTTAAGAGAAATGGCAATCGGCATGCAATAGACA 2725
QY 2169 TAGCGCTGATGGCCCTCAAGATGATGAGCTC-----TCTGATGAAGTTATGTCTCCCC 2222
DB 2726 TTGCAAGATGGCTTGGAAATCTCAGCTTCATGGGACCTTTGAGCTGGAGCATCTTC 2785
QY 2223 ATGGAACCTATCAGATCGAATTTGGAGTGGACCTGATCATGATGATTTTGGCTGGCTG 2282
DB 2786 CTGCGCTCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2845
QY 2283 TTGAGTTAAATGCCCGCTTACTGTCTTTTGGAAACATGTCACCTCTGCTTAAACAAAT 2342
DB 2846 TGGAAATCAGATGCTGCTGTTATGCTATTTGGAGATGATGATGATGATGATGATG 2905
QY 2343 TTGAGTCTGAGTGTACCAAGAAATCAATGTCAG 2379
DB 2906 TGGAAATCACTGGCTCCCTTTGAGAAATTCACGTGAG 2942

RESULT 5

US-08-908-643C-82
; Sequence 82, Application US/080908643C
; Patent No. 6120995

GENERAL INFORMATION:

APPLICANT: Waldman, Scott A.
Pearlman, Joshua M.
Barber, Michael T.
Schultz, Stephanie
Parkinson, Scott J.

TITLE OF INVENTION: COMPOSITIONS THAT SPECIFICALLY BIND TO
COLORECTAL CANCER CELLS AND METHODS OF
USING THE SAME

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 6120995ris LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch disk, 1.44 Mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/908,643C
FILING DATE: 07-Aug-1997
CLASSIFICATION: N/A

PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mark Deluca
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TUTU-2209
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 3787
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-08-908-643C-82

Query Match 4.0%; Score 121.4; DB 3; Length 3787;
Best Local Similarity 53.2%; Pred. No. 8.8e-23;
Matches 307; Conservative 0; Mismatches 261; Indels 9; Gaps 2;

QY 1812 GGAAGCTGAAGGCTACCTTGGAGCAAGCCACCAAGCCCTGGAGGAGGAGAGAGAAAGA 1871
DB 2435 GAAACCTGGACATCTGGTAGAGAAAGGACACAGCTGTACAAGGCGAGAGGACAGGG 2494
QY 1872 CAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTGTCTCAGCAGCTGTGCAAGGGC 1931
DB 2495 CTGACAGATTAACCTTATGCTTCCAAAGGCTAGTGGTAAAGTCTCTCAAGGAGAAAG 2554
QY 1932 AAGTTGTGCAAGCAAGAACTTCAATATGTCACCATGCTCTTCTCAGACATCTGTTGGGT 1991
DB 2555 GCTTTGGAGCGGAACTATATGAGGAAGTTACAATCTACTTCACTGACATTTAGGTT 2614
QY 1992 TCAGTGCATCTCTCCAGTGTCTACCGCTGAGGTGATGATGATGATGATGATGATGATG 2051
DB 2615 TCACTACTATCTGCAATATACAGCACCCCATGGAAGTGGGACATGCTTAAAGCATCT 2674
QY 2052 ACAGCTGCTTCGACAGCAGTGTGGAGAGCTGTCTACAAGGTGGAGACCATTTGGCG 2111
DB 2675 ATAAGAGTTTGACACATTTGTTGATCATCATGATGTCTACAAGGTGGAAACCATCGGTG 2734
QY 2112 ATGCTATTGTGTAGCTGGGGATTAC---ACAAAGAGATGATGATGATGATGATGATGATG 2168
DB 2735 ATGCGTACATGTTGGCTAGTGGTTGCTTAAGAGAAATGGCAATCGGCATGCAATAGACA 2794
QY 2169 TAGCGCTGATGGCCCTGMAAGATGATGAGCTC-----TCTGATGAAGTTATGCTCTCCC 2222
DB 2795 TTGCCAAGATGGCTTTGGAATCTCAGCTTCATGGGAGCTTTGAGCTGGAGCATCTTC 2854
QY 2223 ATGGAAGCACTATCAAGATGCGAATTTGGACTGCACTCTGGATCAGTGTGCTGGGCTCG 2282
DB 2855 CTGGCTCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2914
QY 2283 TTGAGTTAAATGCCCGCTTACTGTCTTTTGGAAACATGTCACCTCTGCTGGCTAACAAAT 2342
DB 2915 TGGGAATCAAGATGCTGCTGTTATGCTATTTGGAGATGATGATGATGATGATGATGATG 2974
QY 2343 TTGAGTCTGAGTGTACCAAGAAATCAATGTCAG 2379
DB 2975 TGGAAATCACTGGCTCCCTTTGAGAAATTCACGTGAG 3011

RESULT 6

US-09-473-716-1
; Sequence 1, Application US/09473716
; Patent No. 6436672

GENERAL INFORMATION:

APPLICANT: Tomlinson, James
APPLICANT: Cor Therapeutics, Inc.
TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A HUMAN ADENYLAL
TITLE OF INVENTION: CYCLASE

FILE REFERENCE: 44481-5027-01-US
CURRENT APPLICATION NUMBER: US/09/473,716
CURRENT FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: PCT/US98/13540
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/070,901
PRIOR FILING DATE: 1997-07-01
PRIOR APPLICATION NUMBER: 08/886,362
PRIOR FILING DATE: 1997-07-01
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4523
TYPE: DNA
ORGANISM: human type V adenyllyl cyclase
FEATURE:
NAME/KEY: CDS
LOCATION: (139)..(3921)
US-09-473-716-1

Query Match 3.0%; Score 91.4; DB 4; Length 4523;
Best Local Similarity 49.9%; Pred. No. 1.2e-14;
Matches 230; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

QY 1945 CAAGAAGTTCAGTATGTCACCATGCTCTCTTCACACATCGTTGGTTCACATGCTG 2004
DB 1524 CCAGAAACATGACAGAGGAGTCTCTGTTGCTGACATCGAGGCTTCACACGCTGCG 1583
QY 2005 CTCCAGTGTCTCAGGCTGACGATCATCACCATGCTCAATGCTCAATGCTGACATCGGCTGCA 2064
DB 1584 GTCCAGTGTCTCAGGCTGACGATCATCACCATGCTCAATGCTGACATCGGCTGCA 1643
QY 2065 CCAGAGTGTGAGAGTGGATGCTCAAGGTGGAGACCATGTCGATGCTGATGCTGATGCTG 2124
DB 1644 CAAGTGTGCTGAGAGTGGATGCTCAAGGTGGAGACCATGTCGATGCTGATGCTGATGCTG 1703
QY 2125 AGCTGGGATGTACAAAGAGTGGATGCTCAAGGTGGAGACCATGTCGATGCTGATGCTGATGCTG 2184
DB 1704 CTTCTGGGCTGCTGAGAGGCTGACCATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1763
QY 2185 GAAGATGAGGAGTCTCTGATGAAGTATGCTCTCCCATGAGAGACCATGTCAGATGCTG 2244
DB 1764 GGACATGATGAGGCTGCTGCTGCTGGGAGGTGACAGGCTGACAGTGAACATCGG 1823
QY 2245 AATTGAGTGTGAGTGTGATGCTGCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2304
DB 1824 TGTGGGATTCACAGCGGCGAGTACACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1883
QY 2305 CTGCTCTTTTGGAAACAATGCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2364
DB 1884 CGAGTGTGCTGAGGATGCTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1943
QY 2365 AAAATCAATGTCAGCCCAACAACTTACAGATTAATCTCAAG 2405
DB 1944 ACGCATCCACATCACCAGGCTTACACTCACTTACCTGATG 1984

RESULT 7
US-09-412-210-2
Sequence 2, Application US/09412210
Patent No. 6403358
GENERAL INFORMATION:
APPLICANT: Kapeller-Libermann, Rosana
TITLE OF INVENTION: 21529, A NOVEL ADENYLYL CYCLASE
FILE REFERENCE: 5800-47
CURRENT APPLICATION NUMBER: US/09/412,210
CURRENT FILING DATE: 1999-10-05
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3518
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: misc_feature
LOCATION: (0)...(0)
OTHER INFORMATION: 21529 adenylate cyclase
NAME/KEY: CDS
LOCATION: (247)...(3480)
US-09-412-210-2

Query Match 3.0%; Score 90.8; DB 4; Length 3518;
Best Local Similarity 50.5%; Pred. No. 1.6e-14;
Matches 221; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 1941 AAGCCAAAGTTCAGTAAATGTCACCATGCTCTCTTCAGACATCGTTGGTTCACATGCA 2000
DB 1040 ATGTCAAGAGGACCAAGGAGTCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1099
QY 2001 TCTGCTCCAGTGTCTCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 2060
DB 1100 TGGCCAGGAGTGTCTCCCTTAAGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1159
QY 2061 TCGACCAAGTGTGAGGAGTGTCTTACAAAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTG 2120
DB 1160 TCGACCAAGTGTGAGGAGTGTCTTACAAAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTG 1219
QY 2121 GTGTAGTGTGAGGAGTGTCTTACAAAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGG 2180
DB 1220 ACTGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1279
QY 2181 CCTCAAGATGATGAGGCTCTCTGATGAAGTGTATGCTCTCCCATGAGGAGTGTGAGGAGTGT 2240
DB 1280 GCTTGAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1339
QY 2241 TCGAATGTGAGTGTGAGGAGTGTCTTACAAAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGT 2300
DB 1340 TCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1399
QY 2301 GTTACTGTCTTGTGAGGAGTGTCTTACAAAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGT 2360
DB 1400 AGTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1459
QY 2361 CACGAAATCAATGTCA 2378
DB 1460 CAGGCGAGTGCACATCA 1477

RESULT 8
US-08-726-214-9
Sequence 9, Application US/08726214
Patent No. 6107076
GENERAL INFORMATION:
APPLICANT: Tang, Wei-Jen
TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,214
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,498


```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 4008 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-726-214-3

Query Match          2.7%; Score 80.6; DB 3; Length 4008;
Best Local Similarity 49.2%; Pred. No. 9.7e-12;
Matches 212; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 1931 CAAGTTGTGCAAGCCCAAGAGTTTCAGTAATGTCACCATGCTCTTCACAGACATCGTTGGG 1990
Db 901 CACAATCTGTATGTCAACGACACACCAACGTCGAGCATATATACCTGACATTTGTGGC 960
QY 1991 TTCACGCGCTTGCAGCGATTGCTCCCTGGGAACTGGTCCACATGCTGTAATGAATC 1020
Db 961 TTCACCGCGCTTGCAGCGATTGCTCCCTGGGAACTGGTCCACATGCTGTAATGAATC 1020
QY 2051 TACACTCGCTTCGACACGAGTGTGGAGAGCTGGATGTCTACAAGGTGGAGACCATTTGGC 2110
Db 1021 TTTGGGAAGTTTGATCAAAATAGCAAGAGAGATGAATGATGAGATTTAAAATTTTAGGA 1080
QY 2111 GATGCGCTATTGTAGCTGGGGATTACACAAGAGAGTGATGATGATGCTGTTTCAGATA 2170
Db 1081 GACTGCTATTACTGTCTTCCGGCTCCCTATATCACTCCCTAACCATGCCAAGAACTGT 1140
QY 2171 GCGCTGATGCGCTTGAAGATGATGGAGCTCTCTGATGAAGTTATGTCTCCCATCGGAA 2230
Db 1141 GTGAAATGGATGGATATGTGCGAAGCCATAAAGAAAGTGAAGGATGCTACCGGAGTT 1200
QY 2231 CCTATCAAGATGGAATTTGAGTGTGCTGATGATGATGATGATGATGATGATGATGATG 2290
Db 1201 GATATCAACATGCGTGTAGAGTGCATTTGGGAACTGCTCTGTGTGTGTGTGTGTGTCT 1260
QY 2291 AAAATCCCGCTTACTGTCTTTTGGAAACAATGTCACTGCTGCTGCTGCTGCTGCTGCTG 2350
Db 1261 CAGAAGTGGCAGTATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCT 1320
QY 2351 TGCAGTGTACC 2361
Db 1321 GGAGGAGTCCC 1331

RESULT 11
PCT-US95-11808-5
; Sequence 5, Application PC/TUS9511808
; GENERAL INFORMATION:
; APPLICANT: Iyengar, Srinivas Ravi V.
; TITLE OF INVENTION: MUTANT ACTIVATED GSALPHA AND
; TITLE OF INVENTION: ADENYLIC
; TITLE OF INVENTION: CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue and
; ADDRESSEE: Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11808
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/307,896
; FILING DATE: 16-SEP-1994

; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Richard S.
; REGISTRATION NUMBER: 26,154
; REFERENCE/DOCKET NUMBER: 29970 165/28755
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 408-2500
; TELEFAX: (212) 765-2519
; TELEX: 650 6111063
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4008 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US95-11808-5

Query Match          2.7%; Score 80.6; DB 5; Length 4008;
Best Local Similarity 49.2%; Pred. No. 9.7e-12;
Matches 212; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 1931 CAAGTTGTGCAAGCCCAAGAGTTTCAGTAATGTCACCATGCTCTTCACAGACATCGTTGGG 1990
Db 901 CACAATCTGTATGTCAACGACACACCAACGTCGAGCATATATACCTGACATTTGTGGC 960
QY 1991 TTCACGCGCTTGCAGCGATTGCTCCCTGGGAACTGGTCCACATGCTGTAATGAATC 1020
Db 961 TTCACCGCGCTTGCAGCGATTGCTCCCTGGGAACTGGTCCACATGCTGTAATGAATC 1020
QY 2051 TACACTCGCTTCGACACGAGTGTGGAGAGCTGGATGTCTACAAGGTGGAGACCATTTGGC 2110
Db 1021 TTTGGGAAGTTTGATCAAAATAGCAAGAGAGATGAATGATGAGATTTAAAATTTTAGGA 1080
QY 2111 GATGCGCTATTGTAGCTGGGGATTACACAAGAGAGTGATGATGATGCTGTTTCAGATA 2170
Db 1081 GACTGCTATTACTGTCTTCCGGCTCCCTATATCACTCCCTAACCATGCCAAGAACTGT 1140
QY 2171 GCGCTGATGCGCTTGAAGATGATGGAGCTCTCTGATGAAGTTATGTCTCCCATCGGAA 2230
Db 1141 GTGAAATGGATGGATATGTGCGAAGCCATAAAGAAAGTGAAGGATGCTACCGGAGTT 1200
QY 2231 CCTATCAAGATGGAATTTGAGTGTGCTGATGATGATGATGATGATGATGATGATGATG 2290
Db 1201 GATATCAACATGCGTGTAGAGTGCATTTGGGAACTGCTCTGTGTGTGTGTGTGTGTCT 1260
QY 2291 AAAATCCCGCTTACTGTCTTTTGGAAACAATGTCACTGCTGCTGCTGCTGCTGCTGCTG 2350
Db 1261 CAGAAGTGGCAGTATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCT 1320
QY 2351 TGCAGTGTACC 2361
Db 1321 GGAGGAGTCCC 1331

RESULT 12
US-08-726-214-7
; Sequence 7, Application US/08726214
; Patent No. 6107076
; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gliman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLIC CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
```



```
QY 2127 CTGGGGATTACAAAGAGAGTGATCTATCTGCTGTTTCAGATAGCGCTGATGCCCTGA 2186
Db      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
662 TGTACGGGCTGCCGGAGGCCCGGCCGACCATGCCACTGCTGCTGTGGAGATGGGGTAG 721
QY 2187 AGATGATGGGAGCTCTCTGATGAAGTTATGTCTCCCCATGGAGAACCTATCAAGATGCGAA 2246
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
722 ACATGATGTGAGGCCATCTCCTGCTAGCGTGAGGTGACAGGTGTGAATGTGAACATGCGCG 781
QY 2247 TTGGACTGCACCTCTGGATCAGTTTCTGCTGGCGTCGTTGGAGTTAAATGCCCGTTACT 2306
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
782 TGGGCATCCACAGGGCGCGTGCACCTGCGCGCTCCTTGGCTTGGGAATGGCAGTTGG 841
QY 2307 GTCTTTTGGAAACAATGTCTCCTGCTTAACAAATTTGAGTCTGCGAGTGTACCACGAA 2366
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
842 ATGTGTGGTCCAATGATGTGACCTGGCCCAACCATGGAAGCAGGAAGCCGGGCTGGCC 901
QY 2367 AATCATGTGACCCCAACAC 2388
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
902 GCATCCACATCACTCGGGCAAC 923
```

Search completed: July 1, 2003, 17:29:22
Job time : 109.089 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 16:02:43 ; Search time 286.696 Seconds
(without alignments)
15611.019 Million cell updates/sec

Title: US-09-762-767A-1

Perfect score: 3015
Sequence: 1 cccttatggcgattggcggc.....aaaaaaaaaaaaaaaaaaaaa 3015

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PTCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PTCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2964	98.3	3017	9	US-10-205-823-161
2	2540	84.2	2850	9	US-10-205-823-163
3	2540	84.2	2871	9	US-10-205-823-164
4	1552	51.5	2430	9	US-09-952-213D-1
c 5	714.6	23.7	177556	9	US-09-952-213D-6
6	483.6	16.0	7697	9	US-09-952-213D-4
7	435	14.4	1037	9	US-10-198-846-13184
8	410.6	13.6	634	9	US-10-198-846-9332
9	308.4	10.2	852	9	US-10-198-846-6205
c 10	176.4	5.9	2335	10	US-09-917-800A-1569
11	127	4.2	7697	9	US-09-952-213D-4
12	121.4	4.0	3745	9	US-10-157-031-17
13	121.4	4.0	3787	10	US-09-819-249-1
14	91.4	3.0	4523	9	US-10-175-158-1
15	90.8	3.0	3518	9	US-10-121-911-2
16	77.6	2.6	3582	10	US-09-750-240-12
17	75.6	2.5	1812	10	US-09-750-240-3
18	75.6	2.5	3549	10	US-09-750-240-5
19	75.6	2.5	3552	10	US-09-750-240-10

Sequence 1, Appli	20	75.6	2.5	4942	9	US-10-201-000-1
Sequence 352, App	21	72	2.4	2601	10	US-09-925-297-352
Sequence 1, Appli	22	70.6	2.3	6196	9	US-10-282-942-1
Sequence 5, Appli	23	67	2.2	15093	9	US-09-952-213D-5
Sequence 6, Appli	24	65.4	2.2	177556	9	US-09-952-213D-6
Sequence 1, Appli	25	62.4	2.1	4985	12	US-10-071-223-1
Sequence 1, Appli	26	61.6	2.0	4473	10	US-09-751-100B-1
Sequence 98, Appli	27	60.8	2.0	5515	10	US-09-751-100B-98
Sequence 182, App	28	59.8	2.0	330	9	US-09-764-868-182
Sequence 71, Appli	29	59.8	2.0	330	9	US-09-989-442-71
Sequence 445, App	30	59.8	2.0	330	9	US-10-091-504-445
Sequence 445, App	31	59.8	2.0	330	10	US-09-764-869-445
Sequence 6274, Ap	c 32	55.2	1.8	418	9	US-09-796-692-6274
Sequence 6274, Ap	c 33	55.2	1.8	418	9	US-10-040-862-6274
Sequence 1351, Ap	34	51.4	1.7	11881	9	US-09-764-868-1351
Sequence 1353, Ap	35	51.4	1.7	11881	9	US-09-764-868-1353
Sequence 2124, Ap	36	51.4	1.7	11881	9	US-10-091-504-2124
Sequence 2124, Ap	37	51.4	1.7	11881	10	US-09-764-869-2124
Sequence 2124, Ap	38	49.2	1.6	459	9	US-09-918-995-32338
Sequence 32338, A	39	49	1.6	778	9	US-09-774-639-30
Sequence 30, Appli	40	49	1.6	778	9	US-09-969-730-38
Sequence 38, Appli	41	48.8	1.6	1073	9	US-10-063-547-95
Sequence 95, Appli	42	48.8	1.6	1073	9	US-10-174-590-305
Sequence 305, App	43	48.8	1.6	1073	9	US-10-176-758-305
Sequence 305, App	44	48.8	1.6	1073	9	US-10-063-616-95
Sequence 95, Appli	45	48.8	1.6	1073	9	US-10-175-737-305

ALIGNMENTS

RESULT 1

US-10-205-823-161
; Sequence 161, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 3017
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-161
Query Match 98.3%; Score 2964; DB 9; Length 3017;
Best Local Similarity 99.8%; Pred. No. 0;

Db 874 AAACAGCTTTCAGTACCTTCTGAACACAGACAGCCATTCGCAAGACGAGGAAAAGGGG 933
QY 1072 CAGGCTTGAGACGCGCTCCATCTTATGCTCGATGAAGAGGATGATTTTCTACATGTTTA 1131
Db 934 CAGGCTTGAGACGCGCTCCATCTTATGCTCGATGAAGAGGATGATTTTCTACATGTTTA 993
QY 1132 CTACTCTCTCCCTTAAGAGAACCACTCCCTGATTTCTCCGGGCATCAATAAGGAGCTGC 1191
Db 994 CTACTCTCTCCCTTAAGAGAACCACTCCCTGATTTCTCCGGGCATCAATAAGGAGCTGC 1053
QY 1192 TCAGTATATATGAACGGAAGTGAAGTGTCTTAATGCTCCCTGCTTCCATATGA 1251
Db 1054 TCAGTATATATGAACGGAAGTGAAGTGTCTTAATGCTCCCTGCTTCCATATGA 1113
QY 1252 TTGAGGAGTGTGTGAATCAGCCCTACTTGTGTACTCCGTTCCATGAAAGACACAA 1311
Db 1114 TTGAGGAGTGTGTGAATCAGCCCTACTTGTGTACTCCGTTCCATGAAAGACACAA 1173
QY 1312 GCCATCCCTGTCCCGACGAAACCCAGTCCTGCTGGTATGCCACATGCTTATCTG 1371
Db 1174 GCCATCCCTGTCCCGACGAAACCCAGTCCTGCTGGTATGCCACATGCTTATCTG 1233
QY 1372 CAAGACATTTCCATTCCTATTCATGTTGACAAAGATATGCAATTTCTGCAATTTGCAA 1431
Db 1234 CAAGACATTTCCATTCCTATTCATGTTGACAAAGATATGCAATTTCTGCAATTTGCAA 1293
QY 1432 TGGCATCAGAGGCTGTATCAACAGAGAGACTTTCAAGGAAGGCTTAATTTGAAGAATA 1491
Db 1294 TGGCATCAGAGGCTGTATCAACAGAGAGACTTTCAAGGAAGGCTTAATTTGAAGAATA 1353
QY 1492 CTTTGAATTTCTGACTCAAAATCAACAGAGAGCTTTAGCGGATCATGACTATGTTGAA 1551
Db 1354 CTTTGAATTTCTGACTCAAAATCAACAGAGAGCTTTAGCGGATCATGACTATGTTGAA 1413
QY 1552 TATCAGTGTGCTAGAGTGAAGAGTGGCAACTCTGTGAAGAACTTCAAGGT 1611
Db 1414 TATCAGTGTGCTAGAGTGAAGAGTGGCAACTCTGTGAAGAACTTCAAGGT 1473
QY 1612 TATGAGCTCAAGGCCAATGATCTACATTTGTAATCCAGTGCATCTGTTTGGG 1671
Db 1474 TATGAGCTCAAGGCCAATGATCTACATTTGTAATCCAGTGCATCTGTTTGGG 1533
QY 1672 GTCACCTGTGTGACAGATAGAGATTTTACAGGACGAGGCTTACCTCTCAGACAT 1731
Db 1534 GTCACCTGTGTGACAGATAGAGATTTTACAGGACGAGGCTTACCTCTCAGACAT 1593
QY 1732 CCCAATTCACATGCACTGAGGATGTGCTTAAATAGGGGAACAAGCCGAGCTCAAGA 1791
Db 1594 CCCAATTCACATGCACTGAGGATGTGCTTAAATAGGGGAACAAGCCGAGCTCAAGA 1653
QY 1792 TGGCCTGAAGAGAGCTGGGGAAGCTGAAGGCTTACCTTTGAGCAAGCCCAAGCCCT 1851
Db 1654 TGGCCTGAAGAGAGCTGGGGAAGCTGAAGGCTTACCTTTGAGCAAGCCCAAGCCCT 1713
QY 1852 GGAGGAGGAGAGAGAGAGAGTACCTTCTGTGCTCCATATTTCCCTGTGAGGTGC 1911
Db 1714 GGAGGAGGAGAGAGAGAGAGTACCTTCTGTGCTCCATATTTCCCTGTGAGGTGC 1773
QY 1912 TCAGCAGCTGTGCAAGGCAAGTGTGCAAGGCAAGAGTTCAGTAAATGTCACCATGCT 1971
Db 1774 TCAGCAGCTGTGCAAGGCAAGTGTGCAAGGCAAGAGTTCAGTAAATGTCACCATGCT 1833
QY 1972 CTTCTCAGACATCGTTGGGTTCACTGCCATCTGCTCCAGTGTCTACCGCTGAGGTCAT 2031
Db 1834 CTTCTCAGACATCGTTGGGTTCACTGCCATCTGCTCCAGTGTCTACCGCTGAGGTCAT 1893
QY 2032 CACCATCTCAATGCACTGTACACTCGCTCGACAGCAGTGTGGAGAGTGTGATGCTA 2091
Db 1894 CACCATCTCAATGCACTGTACACTCGCTCGACAGCAGTGTGGAGAGTGTGATGCTA 1953
QY 2092 CAAGGTGGAGACCATTCGGCATGCTATTTGTAGCTGGGGGATTAACAAAGAGAGTGA 2151

Db 1954 CAAGGTGGAGACCATTCGGCATGCGCTATTTGTAGCTGGGGATTAACAAAGAGTGA 2013
QY 2152 TACTCATGCTGTTTCAGATAGCGCTGATGCGCTGAAGATGATGAGAGTCTCTGTATGAAGT 2211
Db 2014 TACTCATGCTGTTTCAGATAGCGCTGATGCGCTGAAGATGATGAGAGTCTCTGTATGAAGT 2073
QY 2212 TATGCTCTCCCATGGAGAACCTATCAAGATGCGGAATTTGGACTGCACTCTCGATCAGTTTT 2271
Db 2074 TATGCTCTCCCATGGAGAACCTATCAAGATGCGGAATTTGGACTGCACTCTCGATCAGTTTT 2133
QY 2272 TGCTGGCGTCTGTTGGAGCTTAAATGCGCCGTTACTGTCTTTTGGAAACATGTCACCTCT 2331
Db 2134 TGCTGGCGTCTGTTGGAGCTTAAATGCGCCGTTACTGTCTTTTGGAAACATGTCACCTCT 2193
QY 2332 GGGTAAACAAATTTGAGTCTGAGTGTACACGAAATCAATGTCAGCCCAACAACTTA 2391
Db 2194 GGGTAAACAAATTTGAGTCTGAGTGTACACGAAATCAATGTCAGCCCAACAACTTA 2253
QY 2392 CAGATTAATCAAGACTGTCTGCTGTTTTCGCTGTTTACCCCTCGATCAAGGAGGAACTTCC 2451
Db 2254 CAGATTAATCAAGACTGTCTGCTGTTTTCGCTGTTTACCCCTCGATCAAGGAGGAACTTCC 2313
QY 2452 ACCAACTCTCCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGCTTACCACCAAGG 2511
Db 2314 ACCAACTCTCCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGCTTACCACCAAGG 2373
QY 2512 AACAACTCAAAACCATGCTTCCAAAGAAAGATGTGAAAGATGGCAATGCCAATTTTTT 2571
Db 2374 AACAACTCAAAACCATGCTTCCAAAGAAAGATGTGAAAGATGGCAATGCCAATTTTTT 2433
QY 2572 AGGCAAGCATCAGGAATAGATTAGCAACCTATATACCTATTTATAGTCTTTGGGGTTT 2631
Db 2434 AGGCAAGCATCAGGAATAGATTAGCAACCTATATACCTATTTATAGTCTTTGGGGTTT 2493
QY 2632 GACTCATTTGAAGTGTCTAGAGCTCTGGAAGCACTTTAGGGATGTAGATGGCTAACAA 2691
Db 2494 GACTCATTTGAAGTGTCTAGAGCTCTGGAAGCACTTTAGGGATGTAGATGGCTAACAA 2553
QY 2692 GCAGTATTAATAATTTCAAGGACCAAGTCACAATCTTTCTCTCTTTTAAACATGACAAA --- 2748
Db 2554 GCAGTATTAATAATTTCAAGGACCAAGTCACAATCTTTCTCTCTTTTAAACATGACAAA 2613
QY 2749 -ATGCTACTACTTCAGTACTTCAGCTCTTCAAG-AAAAAANAACCTTTAAAAAGCTA 2806
Db 2614 TATGCTACTACTTCAGTACTTCAGCTCTTCAAGAAAAAANAACCTTTAAAAAGCTA 2673
QY 2807 CTTTGTGGGAGTATTTCTATTTATATAACAGCACTTACTACCTGTACTCAAAATTCAGC 2866
Db 2674 CTTTGTGGGAGTATTTCTATTTATATACAGCACTTACTACCTGTACTCAAAATTCAGC 2733
QY 2867 ACCTGTGACATATATCAGATTAATTTAGTCAATTTGTACAACTGATGGAGTACCTGCAA 2926
Db 2734 ACCTGTGACATATATCAGATTAATTTAGTCAATTTGTACAACTGATGGAGTACCTGCAA 2793
QY 2927 TCTCATATCTGCTGGGATGCCATGCTTATTAAGTGTGTTGTGATAGT 2976
Db 2794 TCTCATATCTGCTGGGATGCCATGCTTATTAAGTGTGTTGTGATAGT 2843

RESULT 3

US-10-205-823-164

; Sequence 164, Application US/10205823

; Publication No. US20030108963A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Monahan, John E.

; APPLICANT: Endege, Wilson O.

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Gorbacheva, Bella

; APPLICANT: Hoersch, Sebastian

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Wonsay, Angela M.

; APPLICANT: Glatt, Karen

```

APPLICANT: Zhao, Xumei
APPLICANT: Anderson, Dustin
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
FILE REFERENCE: MRI-044
CURRENT APPLICATION NUMBER: US/10/205,823
CURRENT FILING DATE: 2002-07-25
PRIOR APPLICATION NUMBER: 60/307,982
PRIOR FILING DATE: 2001-07-25
PRIOR APPLICATION NUMBER: 60/314,356
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/325,020
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/341,746
PRIOR FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/362,158
PRIOR FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 164
LENGTH: 2871
TYPE: DNA
ORGANISM: Homo sapiens
US-10-205-823-164

```

Query Match	84.2%	Score 2540;	DB 9;	Length 2871;
Best Local Similarity	99.08;	Pred. No. 0;		
Matches 2565;	Conservative	0;	Mismatches	0; Indels 5; Gaps 2;
QY	412	ACATCCAGTACCAGTGTCCITGAATTGATAGTGGCTTCTGTTTGTCACTCTCATATAA	471	
Db	295	ACATCCAGTACCAGTGTCCITGAATTGATAGTGGCTTCTGTTTGTCACTCTCATATAA	354	
QY	472	GAATCAGAGCTCATCAGGAGGATCGACGAGGTAAAGACACCAACACCATGTTCTG	531	
Db	355	GAATCAGAGCTCATCAGGAGGATCGACGAGGTAAAGACACCAACACCATGTTCTG	414	
QY	532	CACGAAGCTCAAGGATGTCGAAGATCATCAGAGAGTGTCCITTTCTCTTACTGCACACG	591	
Db	415	CACGAAGCTCAAGGATGTCGAAGATCATCAGAGAGTGTCCITTTCTCTTACTGCACACG	474	
QY	592	TCAAGTCTCTAACGAGTCTTCAGAGAGGCGAGAGAGTCTAGAGAGTGCAGAAAGCAAC	651	
Db	475	TCAAGTCTCTAACGAGTCTTCAGAGAGGCGAGAGAGTCTAGAGAGTGCAGAAAGCAAC	534	
QY	652	CGTGCCCATCTGTCAAGACATTCCTGAGAAGACATACAGAAGTCTTCTCTCAAGAAA	711	
Db	535	CGTGCCCATCTGTCAAGACATTCCTGAGAAGACATACAGAAGTCTTCTCTCAAGAAA	594	
QY	712	AACAGTCGGAGCGGAGTCTATCTCACACTTTGGCAGAGAGTATTTGCAAACTGATTTT	771	
Db	595	AACAGTCGGAGCGGAGTCTATCTCACACTTTGGCAGAGAGTATTTGCAAACTGATTTT	654	
QY	772	CCCAGAGTTTGAAGGCTGAATGTTGCACATTCAGAGACATTTGGCAAGACAAAATAA	831	
Db	655	CCCAGAGTTTGAAGGCTGAATGTTGCACATTCAGAGACATTTGGCAAGACAAAATAA	714	
QY	832	AGAAAGCAGGAAATCTTTGGAAGAGAGAGACTTTGAAAAACAATTTGCAGAGCAAGCAGT	891	
Db	715	AGAAAGCAGGAAATCTTTGGAAGAGAGAGACTTTGAAAAACAATTTGCAGAGCAAGCAGT	774	
QY	892	TGCAGCAGGAGTTCCAGTGGAGGTATATCAAGAATCTCTTGGTGAAGAGGTTTTTAAAT	951	
Db	775	TGCAGCAGGAGTTCCAGTGGAGGTATATCAAGAATCTCTTGGTGAAGAGGTTTTTAAAT	834	
QY	952	ATGTTACGAGGAGATGAAAAACATTCCTTGGGGTGGTTGGAGCACCCCTTAAGATTTTTT	1011	
Db	835	ATGTTACGAGGAGATGAAAAACATTCCTTGGGGTGGTTGGAGCACCCCTTAAGATTTTTT	894	
QY	1012	AAACAGGTTTCAGTACCCCTTCGTAAACAGAGCACCATTGCCAAGAGCAGGAAAAAGGGG	1071	
Db	895	AAACAGGTTTCAGTACCCCTTCGTAAACAGAGCACCATTGCCAAGAGCAGGAAAAAGGGG	954	

```
QY 2152 TACTCATGCTGTTTCAGATAGCGGTGATGGCCCTGAAGATGATGAGGCTCTCTGATGAAGT 2211
Db 2035 TACTCATGCTGTTTCAGATAGCGGTGATGGCCCTGAAGATGATGAGGCTCTCTGATGAAGT 2094
QY 2212 TATGTCCTCCCATGGAGAACCTATCAAGATGCGAATTTGGACTGCACTCTCGATCAGTTTTT 2271
Db 2095 TATGTCCTCCCATGGAGAACCTATCAAGATGCGAATTTGGACTGCACTCTCGATCAGTTTTT 2154
QY 2272 TGTGGCGCTGTTGGAGTTAAATGCCCCCTTACTGTCTTTTGGAAACAATGTCACCTCT 2331
Db 2155 TGTGGCGCTGTTGGAGTTAAATGCCCCCTTACTGTCTTTTGGAAACAATGTCACCTCT 2214
QY 2332 GGCTAAACAATTTGAGTCTGCGAGTGTACACAGAAATCAATGTGAGCCCAACAACCTTA 2391
Db 2215 GGCTAAACAATTTGAGTCTGCGAGTGTACACAGAAATCAATGTGAGCCCAACAACCTTA 2274
QY 2392 CAGATTACTCAAGAGCTGCTGCTGTTTCGTTTACCCCTCGATCAAGGAGGAACTTCC 2451
Db 2275 CAGATTACTCAAGAGCTGCTGCTGTTTCGTTTACCCCTCGATCAAGGAGGAACTTCC 2334
QY 2452 ACCAAACTTCCCTAGTGAATCCCGGAATCTGCCATTTTCTGGATGCTTTACCAACAAGG 2511
Db 2335 ACCAAACTTCCCTAGTGAATCCCGGAATCTGCCATTTTCTGGATGCTTTACCAACAAGG 2394
QY 2512 AACAACTCAAAACCATGCTTCCAAAGAAGATGTGGAAGATGGCAATGCCAATTTTTT 2571
Db 2395 AACAACTCAAAACCATGCTTCCAAAGAAGATGTGGAAGATGGCAATGCCAATTTTTT 2454
QY 2572 AGCAAGAGCATCAGGAATAGATTAGCAACCTATATACCTATTATTAAGTCTTTGGGTTTT 2631
Db 2455 AGCAAGAGCATCAGGAATAGATTAGCAACCTATATACCTATTATTAAGTCTTTGGGTTTT 2514
QY 2632 GACTCATTTGAAGATGTAGAGCCTCTGAAAGACCTTTAGGGATTTGATGCTTAACAA 2691
Db 2515 GACTCATTTGAAGATGTAGAGCCTCTGAAAGACCTTTAGGGATTTGATGCTTAACAA 2574
QY 2692 GCAGTATTAATTTTCAGGAGCAAGTCAACAATTTCTCTGTTTAAACATGACAAA --- 2748
Db 2575 GCAGTATTAATTTTCAGGAGCAAGTCAACAATTTCTCTGTTTAAACATGACAAA 2634
QY 2749 -ATGTACTACTTCACTGATCTGAGTCTGAAAGACCTTTAGGGATTTGATGCTTAACAA 2806
Db 2635 TATGTACTACTTCACTGATCTGAGTCTTCAAGAAATTTTAAAGCTTAACAA 2694
QY 2807 CTTTGTGGAGGATTTCTATTATATTAACAGACCTTACTGCTGATCTCAAAATTCAGC 2866
Db 2695 CTTTGTGGAGGATTTCTATTATATTAACAGACCTTACTGCTGATCTCAAAATTCAGC 2754
QY 2867 ACCTTGATATATATATATATATATATATATATATATATATATATATATATATATAT 2926
Db 2755 ACCTTGATATATATATATATATATATATATATATATATATATATATATATATATAT 2814
QY 2927 TCTCATATCTGTTGGAGTCCCATGTTTAAAGTGTGTTTGTGATAGT 2976
Db 2815 TCTCATATCTGTTGGAGTCCCATGTTTAAAGTGTGTTTGTGATAGT 2864
```

RESULT 4

```
US-09-952-213D-1
; Sequence 1, Application US/0995213D
; Publication No. US20030096240A1
; GENERAL INFORMATION:
; APPLICANT: MURAD, FERID
; APPLICANT: SHARINA, IRAIDA G.
; APPLICANT: KRUMENACKER, J. S.
; APPLICANT: MARTIN, E.
; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SGC
; FILE REFERENCE: US/2520S
; CURRENT APPLICATION NUMBER: US/09/952,213D
; CURRENT FILING DATE: 2002-08-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
```

```
; LENGTH: 2430
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (286)..(2361)
US-09-952-213D-1

Query Match      51.5%; Score 1552; DB 9; Length 2430;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1810; Conservative 0; Mismatches 380; Indels 10; Gaps 2;

QY 420 GTTACACAGTCTCTGAAATGTAGTGGCTTCTGTTTGTGTCAGTCTCATATAAGAACTACA 479
Db 189 GTCACCTCTGCTCTTGTAGTCAAGTGAAGCAGATCTTTCATCAGTCCACATCAACACCAAGT 248
QY 480 GCTCATCAGAGGAGATGCGAGCAGGCTAAGAGACACCAACACATGTTCTGCACGAGC 539
Db 249 AGTCAGAAGGAAGCCACTGC-----CAAGCTCCAGGAACACCATGTTCTCGAGGAAGT 301
QY 540 TCAAGGATCTCAAGATCAACAGAGAGTGTCTTTCTCTTCTTCTGACACCAAGTCAAGTTC 599
Db 302 TCAAGGATCTCAAGATCAACAGGAGAGTGTCTTTCTCTTCTTCTGACCTGCTCAGGTT 361
QY 600 CTAAACAGTCTTTCAGAGGAGGAGAGTCTCAGAGAGTCTGCAAGCAACCGTGCCTCA 659
Db 362 CTAAAGGAGCAACAGAGGAGTGGCTGAGGCTCTGAGGCTCTGAGGCTTACTCTGCTCA 421
QY 660 TCTGTCAAGACATCTCTGAGAACACATCAACAAAGTCTTCTCAAGAAACCAAGCAGTC 719
Db 422 TCTGCCAGTACTTCTCTGAGAAATGCAAGAGGAGTCTCCCCCAAGAAAGACAAGCC 481
QY 720 GGAGCGAGTCTATCTTTCACACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCACAGT 779
Db 482 GCAACAGAGTCTACCTGACACACCTGGCAGAGAGTATTTGCAAGTCTATCTTCCACAGT 541
QY 780 TTCAACGGCTGATGTTGACATTCAGAGACATTTGCCAAGACACAAATTAAGAAAGCA 839
Db 542 GTGAGGAGTGAACCTTTCACATTCAGAGAACCTTGGCAAGACATTAATAAGAAAGCA 601
QY 840 GGAATCTTTGGAAGAGAGAGTCTTGAAGAAACATTTGCAGAGCAAGAGTTCAGCAG 899
Db 602 GGAATCTTCAGAAAGAGAGACCTTGAAGAAATATTCGCAAGAGAGCAATTCAGCAG 661
QY 900 GAGTTCAGTGGAGGTTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAATATGTTACG 959
Db 662 GTGCCCCAGTGGAGGCTCTCAAGACTCTCTGGGCGAGGAGCTGTCTCAAGATCTGCTATG 721
QY 960 AGGAAGATGAACATCTCTTGGGCTGGTGGAGGCACTTAAAGATTTTAAACAGCT 1019
Db 722 AGGAAGATGAGCACAATTTTGGGCTGGTGGGCGCACCTTGAAGGACTTTCTTAATAGCT 781
QY 1020 TCAGTACCCCTTCTGAACAGAGCAGGCTTCCAAAGAGAGGAGGAGGAGGAGGAGGAG 1079
Db 782 TCAGCAGCTCTTCAGCAGAGCAGGCTTCCCAAGGCGGAGAGGCGGAGGAGGAGGAGTGG 841
QY 1080 AGGACGCTCCATTTCTATGCTTGATAAGAGGATGATTTTCTACATGTTTACTACTTCT 1139
Db 842 AAGATGCTCCATCTTATGCTTGCAAGACCAAGGACTTTCTTAAATGTTTACTACTTCT 901
QY 1140 TCCCTAAGAGAACCACTCTCTGATTTCTCCCGCATCATAAAGGAGCTGCTCAGCTAT 1199
Db 902 TCCCGAAGAGAACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 961
QY 1200 TATATGAAGAGGAGTGAAGTGTCTGTTAATGCTCTCTCTCTCTCTCTCTCTCTCTCT 1259
Db 962 TGTAAGAGGAGGAGGAGTGTCTCTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1021
QY 1260 AGTTTGTGAATCAGCCCTACTTGTGTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1319
Db 1022 AGTTTGTGAACCAAGCCCTATTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1081
QY 1320 TGTCCCCCAGCAAAACCCAGTCTCTGCTGCTGATTTCCACATGCTCTCTCTCTCTCT 1379
```


Db 24584 CCCTGTCCCCAGCAAAACCCAGTCTCGTGGTGAATTCACATCGCTATTCTGCAAGA 24525
QY 1377 CATTTCATTCCTTTCATCTTTGACAAAGATATGACAATCTGCAATTTGGCAATGGCA 1436
Db 24524 CATTTCATTCCTTTCATCTTTGACAAAGATATGACAATCTGCAATTTGGCAATGGCA 24465
QY 1437 TCAGAAGGCTGTATGAACAGGAGAGACTTTTCAAGGAAGCCCTAAATTTTGAAGAATATCTTG 1496
Db 24464 TCAGAAGGCTGTATGAACAGGAGAGACTTTTCAAGGAAGCCCTAAATTTTGAAGAATATCTTG 24405
QY 1497 AATTTCTGACTCCAAAATCAACAGAGCTTTTACGGGATCATGATCTTTGAATATGC 1556
Db 24404 AATTTCTGACTCCAAAATCAACAGAGCTTTTACGGGATCATGATCTTTGAATATGC 24345
QY 1557 AGTTTGTGTACGAGTGAGGAGATGGGACAACCTCTGTGAAGAAATCTTCAAGGGTATGG 1616
Db 24344 AGTTTGTGTACGAGTGAGGAGATGGGACAACCTCTGTGAAGAAATCTTCAAGGGTATGG 24285
QY 1617 A 1617
Db 24284 A 24284

RESULT 6

US-09-952-213D-4
; Sequence 4, Application US/09952213D
; Publication No. US20030096240A1

; GENERAL INFORMATION:

; APPLICANT: MURAD, FERID

; APPLICANT: SHARINA, IRAIDA G.

; APPLICANT: KRUMENACKER, J. S.

; APPLICANT: MARTIN, E.

; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SGC

; FILE REFERENCE: UFSH-252US

; CURRENT APPLICATION NUMBER: US/09/952,213D

; CURRENT FILING DATE: 2002-08-16

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 7697

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: modified_base

; LOCATION: (605)..(6955)

; OTHER INFORMATION: N = A, C, T/U OR G

US-09-952-213D-4

Query Match 16.0%; Score 483.6; DB 9; Length 7697;
Best Local Similarity 78.9%; Pred. No. 7.5e-127;
Matches 584; Conservative 3; Mismatches 152; Indels 1; Gaps 1;

QY 896 GCAGGAGTTCAGTGAGGTTATCAAGAATCTCTGGTGAAGAGGTTTTTAAATATGT 955
Db 2862 GCAGGTGCCCTGAGGCGCTCAAGACTCTGGGCGAGGAGCTGTTCAGAGACTGC 2921
QY 956 TAGGAGGAAGATGAAGAATCTCTGGTGGTGGAGGACCTTAAAGATTTTTTAAAC 1015
Db 2922 TAGGAGGAAGATGAGACATTTTGGCGTGGTGGGCGCACCTTGAGGACTTCTAAT 2981
QY 1016 AGCTTCAGTACCCTTCTGAAACAGAGCAGCCATTTGCCAAGACAGGAAAGGCGAG 1075
Db 2982 AGCTTCAGCAGCTCTCAAGACAGAGCAGCCACTGCCAAGAGGCGGAGCGGCGAG 3041
QY 1076 CTTGAGGACCCCTCCATCTCTATCCCTGGATAAGGAGGATGATTTTCTACATGTTTACTAC 1135
Db 3042 CTGGAAGATGCCCTCCATTTATGCTGGACAGGACGAGGACTTTCTAAATGTTTACTAC 3101
QY 1136 TTCTTCCCTAAGAGAACCACTCCCTGATTTCTCCCGGCATCATAAAGGCGAGCTGCTCAC 1195
Db 3102 TTCTTCCCGAAGAGAACCAAGCGCTTCTCTCCCTGGTATCATTAAGCGGCTGCTGCG 3161

QY 1196 GTATTATATGAACGGAAGTGAAGTGTCTGTTAATGCCTCCCTGCTTCCATAATGATTCG 1255
Db 3162 ATACTGTACGAAGACCGAGGTGGAGGTGCTCCTGATGCTCCCTGCTTCCGAAGTGAAGTGT 3221
QY 1256 AGCGAGTGTGTGAATCAGCCCTACTTGTGTACTCGTTCACATGAAAGCACCAGCCCA 1315
Db 3222 ACCGAGTGTGTGAACAGCCCTACTTGTGTACTCGTTCACATGAAAGCACCAGCCCA 3281
QY 1316 TCCTCTGCTCCCGAGCAACCCAGTCTCGTGGTGAATTCACATGCTGCTTATCTGC-AA 1374
Db 3282 TCCTCTGCTCCCGAGCAACCCAGTCTCGTGGTGAATTCACATGCTGCTTATCTGC-AA 3341
QY 1375 GACATTTCCATTTCCATTTTCATTTTGAACAAAGATATGACAATTTGCGCAATGG 1434
Db 3342 GACTTTCCCGTTTSCATTTTCATGCTGGACCGAGACCTGGCCATCTGCGAGCTGGTACGG 3401
QY 1435 CATCAGAAGGCTGATGAACAGGAGAGACTTTCAAGGAAGCCCTAAATTTTGAAGAACTTT 1494
Db 3402 CATCAGAAGGCTGATGAACAGGAGAGACTTTCAAGGAAGCCCTAAATTTTGAAGAACTTT 3461
QY 1495 TGAATTTCTGACTCCAAAATCAACAGAGCTTTTAGCGGATCATGACTATGTTGAATAT 1554
Db 3462 TGAATTTCTGACTCCAAAATCAACAGAGACTTTTAGCGGATCATGACTATGTTGAATAT 3521
QY 1555 GCAGTTTGTGTACGAGTGAGGAGATGGGACAACCTCTGTGAAGAAATCTTCAAGGGTAT 1614
Db 3522 GCAGTTTGTGTACGAGTGAGGAGATGGGACAACCTCTGTGAAGAAATCTTCAAGGGTAT 3581
QY 1615 GGACCTCAAGGCGCAATGA 1634
Db 3582 GAAATGTAAACGCGGATCA 3601

RESULT 7

US-10-198-846-13184
; Sequence 13184, Application US/10198846
; Publication No. US20030099974A1

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; APPLICANT: Steinmann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF BREAST CANCER

; FILE REFERENCE: MRI-049

; CURRENT APPLICATION NUMBER: US/10/198,846

; CURRENT FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 60/306,220

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13184

; LENGTH: 1037

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1037

; OTHER INFORMATION: n = A, T, C or G

US-10-198-846-13184

Query Match 14.4%; Score 435; DB 9; Length 1037;
Best Local Similarity 98.0%; Pred. No. 1.5e-113;
Matches 483; Conservative 0; Mismatches 5; Indels 5; Gaps 4;

QY 412 ACATCCAGTTACCACTGCTTGAATGATGAGTGGCTTCTGTTTGTGAGTCTCATATAA 471
Db 210 ACATCCAGTTACCACTGCTTGAATGATGAGTGGCTTCTGTTTGTGAGTCTCATATA 269
QY 472 GAACACTACGCTATCAGGAGGATCGCAGCGGTAAGACACCAACCACTGCTCTG 531
Db 270 GAACACTACGCTATCAGGAGGATCGCAGCGGTAAGACACCAACCACTGCTCTG 329

649	Qy	AACCGTGCCCATCTCTGCTCAAGACATTCCTTGAGAACAACATACAAGAAGTCTTCCCTCAAG	708
491	Db	AACCGTGCCCATCTCTGCTCAAGACATTCCTTGAGAACAACATACAAGAAGTCTTCTTAAAG	550
709	Qy	AAAACCCAGT-CCGAGCCGAGTCTATCTTACACTTTGGCA-GAGAGTATTTCGAAACTG	766
551	Db	AAAACCCAGTCCGGAGCCGAGTCTATCTTACACTTTGGCAAGAAAGTATTTCAAACTG	610
767	Qy	ATTTTCCAGAG-TTTGAACCGCTGAA-TGTTGCATTCAGAGAA	809
611	Db	ATTTTCCAGAGTTTTGAACGGCTGAATGTGTTACACTTCANAAA	655

RESULT 10

```

US-09-917-800A-1569
; Sequence 1569, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1569
; LENGTH: 2335
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_012770
US-09-917-800A-1569

```

	Query Match	5.9%	Score 176.4;	DB 10;	Length 2335;
	Best Local Similarity	52.3%;	Pred. No. 2.7e-39;		
	Matches 415;	Conservative 0;	Mismatches 376;	Indels 3;	Gaps 1;
QY	1617	ACCTCAAAGCCAAATGATCTACATCTCTGAATCCAGTCGAATCTCTGTTTTGGGGTCCAC	1676		
Db	1021	AACTCCGGGTGAGATGATCTGGATGAGTCTCTGAGTGCATGATCTTCATGTGTTCCC	1080		
QY	1677	CCTGTGTGGACAGATTAGAAAGATTTTACAGGACGAGGGCTCTACCTCTCAGACATCCCAA	1736		
Db	1081	CAAAAGTCCGACGCTGCAAGAGCTGGAAGAGAGCAAGATGCATCTTCTTGATATCGCTC	1140		
QY	1737	TTCACAATGCATGAGGGATGTGGTCTTATAGGGGACAAAGCCCGAGTCAAGATGGCC	1796		
Db	1141	CGCAGACAGACGACCGAGGATCTCATCTCTCTCAACACAGAGAGGCTGGCAGATGGAGC	1200		
QY	1797	TGAAGAAGAGGCTGGGGAAGCTCAAGAGCTACCCCTTGAGCAAGGCCCAAGAGCCCTGGAGG	1856		

Db	1201	TTGCTCTGCCAATCGGAAAGGAAGGAGGATTGGCTGCTCTTCCAATCACTGTGCCCA	1260
Qy	1857	AGGAGAAAGAAAAAGACAGTAGACCTTCTGTGCTGCCATATTTCCTGTGAGGTTGCTCAGC	1916
Db	1261	TCGAGAAGAAGACAGACAGACCTTCTGTATGCCATGCTGCCTGAACATGTGGCCAAACC	1320
Qy	1917	AGCTGTGGCAAGGGCAAGTTGTCAAGCCAAAGATTCAAGTAATGTCAACATGCTCTTCT	1976
Db	1321	AATCTAAGGAGGCGAAGAGTGGCTGCAGGAGAAATTTGAACATGTACAATCCTTTTCA	1380
Qy	1977	CAGACATCGTTGGTTTCACGTCCCATCTGCTCCCAAGTCTCACCGCTGCAGGTCAACACCA	2036
Db	1381	CGCATGTTGTGACATTTACCAACATCTGTGCCACCTGTGAACCTATCCAAATCGTGAACA	1440
Qy	2037	TGCTCAATGCACCTGTACACTCGCTTCGACACAGCAGTGTGGAGAGCTGGAGTGTCTACAAGG	2096
Db	1441	TGCTGTAATTCAAATGTACTCCAAAGTTTCACAGGTTTAACACAGTGTCCATGATGTCTACAAAG	1500
Qy	2097	TGGAGACCATTTGGCGATGCTTAATTTGTGTAGCTCGGGGATTTACACAAGAGAGTGATATCTC	2156
Db	1501	TAGAAACATAGGGGATGCTTACATGGTGGTGGGTGAGTACCAGTACCGGTGTGAAGCC	1560
Qy	2157	ATGCTGTTTCAGATAGCGCTGATGCCCTGAAGATGATGGAGCTCTCTGATGAAGTTATGT	2216
Db	1561	ATGCTCAAAGAGTCGGCAATTTTGTCTGGGGATGAGAAATTTCTGCAAAAGAAGTGTATGA	1620
Qy	2217	CTCC--CCATGAGAACCTATCAAGATCGGAATTCGACTGCACTCTGGATCAGTTTTCG	2273
Db	1621	ATCCTGTCACTGGGGAACCTATCCGATCAGATGGGAAATCCACATGGACCAGTCTTAG	1680
Qy	2274	CTGGCGTCGTTGGAGTTAAATCCCGTTACTGTCTTTTTGGAAACAATGTCACTCTGG	2333
Db	1681	CAGGTGTTGTGGGAGACAAGATGCCTCGGTACTGCTTTTGTGTGACACTGTAAACACAG	1740
Qy	2334	CTAACAAATTTGAGTCTCTCAGGTACCAGGAATAATCAATGTCAAGCCCAAGCACTTACA	2393
Db	1741	CCTCTAGGATGGAAAGTCACGGGCTTCCCAAGCAAGTGCATCTGAGCCCCACAGCCACA	1800
Qy	2394	GATTACTCAAAAGAC	2407
Db	1801	GAGCCCTGAAAAAC	1814

```

RESULT 11
US-09-952-213D-4/c
; Sequence 4, Application US/09952213D
; Publication No. US20030096240A1
; GENERAL INFORMATION:
; APPLICANT: MURAD, FERID
; APPLICANT: SHARINA, IRAIDA G.
; APPLICANT: KRUMENACKER, J. S.
; APPLICANT: MARTIN, E.
; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SC
; FILE REFERENCE: UTSH:252US
; CURRENT APPLICATION NUMBER: US/09/952,213D
; CURRENT FILING DATE: 2002-08-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 7697
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (605)..(6955)
; OTHER INFORMATION: N = A, C, T/U OR G
US-09-952-213D-4

```

Query Match 4.2%; Score 127; DB 9; Length 7697;
Best Local Similarity 87.4%; Pred. No. 8.7e-25;
Matches 139; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

1797	CGCAGCAGCAGCAGGGATCTCATCTCTCTCAACAGCAGAGGCTGGCAGAGATGGAGC	1200
1797	TGAAGAAGAGGCTGGGGAAGCTGAAGGCTACCTCTTGAGCAAGCCCAAGAGCCCTGGAGG	1856

QY 2238 AGATCGGAATTTGGACTGCACTCTGGATCAGTTTTCCTGGCGTCGTTGGAGTTAAATGC 2297
|||||
Db 6509 AGATCGGAATTTGGACTACATCTGGATCAGTTTTCCTGGAGTTTCGGAGTGAAGATGC 6450
QY 2298 CCGGTTACTGCTCTTTTGGAAACAATGTCACTCTGGCTAACAAATTTGAGTCTCGCAGTG 2357
|||||
Db 6449 CCGGTTATTCCTGTTTGGAAACAATGTCACTCTGGCTAACAAATTTGAATCTCGCAGTG 6390
QY 2358 TACCAGGAATAATCAATGTGAGCCCAACAACTTACAGAT 2396
|||||
Db 6389 TGCTCGGAATAATCAATGTGAGCCCAACCAATACAGGT 6351

RESULT 12

US-10-157-031-17
; Sequence 17, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 3745
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-157-031-17

Query Match 4.0%; Score 121.4; DB 9; Length 3745;
Best Local Similarity 53.2%; Pred. No. 2.1e-23;
Matches 307; Conservative 0; Mismatches 261; Indels 9; Gaps 2;
QY 1812 GGAAGCTGAAGGCTACCTTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAGAAAAGA 1871
|||||
Db 2366 GAAACCTGGAACATCTGGTAGAGGAAGGACACAGCTGTACAGGAGGAGGACAGG 2425
QY 1872 CAGTAGACCTTCTGCTCCATATTTCCCTGTGAGGTGCTCAGCAGCTGTGGCAAGGC 1931
|||||
Db 2426 CTGACAGACTTAATCTTTATGTTCTTCCAGGCTAGTGTAAAGTCTCTGAAGGAGAAAG 2485
QY 1932 AAGTTCTGCAAGCCAAAGAGTTTCAGTAATGTCAACATGCTTCTCAGACATCGTTGGGT 1991
|||||
Db 2486 GCTTTGTGAGCGCGGAACATATATGAGGAAGTTACAAATCTACTTCACTGACATTTGAGTT 2545
QY 1992 TCACTGCCATCTCTCCAGTGTCTACCGCTGCAGGTCTATCACCATGCTCAATGCACTGT 2051
|||||
Db 2546 TCACTACTATCTCAAAATACAGCACCCCAAGCTGAGTGGACATGCTTAATGACATCT 2605
QY 2052 ACACCTGCTTCGACAGCAGTGTGGAGAGCTGTATCTACAGGTGGAGACCATTTGGCG 2111
|||||
Db 2606 ATAAGAGTTTGTACCAATCTTTGATCATCATGATCTCTACAGGTGGAAACCATCGGTG 2665
QY 2112 ATCCCTATTGTGTAGCTGGGGATTAC---ACAAGAGAGTGATATCTCATGCTGTTTCAGA 2168
|||||
Db 2666 ATCCGTACATGTGTGGCTAGTGGTTTGCCTTAAGAGAAATGCAATCGGCATGCAATAGACA 2725
QY 2169 TAGCGGTGATGGCCCTGAAGATGATGGAGTCTC-----TCTGATGAAGTTATGCTCTCCCC 2222
|||||
Db 2726 TTGCCAAGATGGCTTGGAAATCTCAGCTTCATGGGACCTTTGAGCTGGAGCATCTTC 2785
QY 2223 ATGAGAACCTATCAAGATGCGAAATTTGGACTGCACTCTGGATCAGTTTTCCTGGCGTCG 2282
|||||
Db 2786 CTGGCTCCCAATATGATTCGCATTTGGAGTTTCACTCTGGTCCCTGCTGCTGGAGTTG 2845
QY 2283 TTGGAGTTAAATGCCCCGTTACTGCTTTTTCGAAACAATGTCACTCTGGCTGCTTACAAAT 2342
|||||

Db 2846 TGGGAATCAAGATGCCCTCGTTATTGCTATTGAGGATACGGTCAACACAGCCCTCTAGGA 2905
QY 2343 TTGAGTCCCTGCAGTGTACCAAGAAATCAATGTCTAG 2379
|||||
Db 2906 TGGAAATCCACTGSCCTCCCTTTTGAGAAATTCACGTGAG 2942

RESULT 13

US-09-819-249-1
; Sequence 1, Application US/09819249
; Patent No. US20010029019A1
; GENERAL INFORMATION:
; APPLICANT: Waldman, Scott A.
; APPLICANT: Park, Jason
; APPLICANT: Schulz, Stephanie
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting Cancer
; TITLE OF INVENTION: Alimentary Canal Origin
; FILE REFERENCE: TJU2412
; CURRENT APPLICATION NUMBER: US/09/819,249
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,229
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3787
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (118)..(3336)
US-09-819-249-1

Query Match 4.0%; Score 121.4; DB 10; Length 3787;
Best Local Similarity 53.2%; Pred. No. 2.1e-23;
Matches 307; Conservative 0; Mismatches 261; Indels 9; Gaps 2;
QY 1812 GGAAGCTGAAGGCTACCTTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAGAAAAGA 1871
Db 2435 GAAACCTGGAACATCTGGTAGAGGAAGGACACAGCTGTACAAAGGAGGAGGACAGG 2494
QY 1872 CAGTAGACCTTCTGCTCCATATTTCCCTGTGAGGTGCTCAGCAGCTGTGGCAAGGC 1931
Db 2495 CTGACAGACTTAATCTTTATGTTCTTCCAGGCTAGTGTAAAGTCTCTGAAGGAGAAAG 2554
QY 1932 AAGTTCTGCAAGCCAAAGAGTTTCAGTAATGTCAACATGCTTCTCAGACATCGTTGGGT 1991
Db 2555 GCTTTGTGAGCGCGGAACATATATGAGGAAGTTACAAATCTACTTCACTGACATTTGAGTT 2614
QY 1992 TCACTGCCATCTCTCCAGTGTCTACCGCTGCAGGTCTATCACCATGCTCAATGCACTGT 2051
Db 2615 TCACTACTATCTCAAAATACAGCACCCCAAGTGGTGGACATGCTTAAATGACATCT 2674
QY 2052 ACACCTGCTTCGACAGCAGTGTGGAGAGCTGTATCTACAGGTGGAGACCATTTGGCG 2111
Db 2675 ATAAGAGTTTGTACCAATCTTTGATCATCATGATCTCTACAAAGGTGGAAACCATCGGTG 2734
QY 2112 ATCCCTATTGTGTAGCTGGGGATTAC---ACAAGAGAGTGATATCTCATGCTGTTTCAGA 2168
Db 2735 ATCCGTACATGTGTGGCTAGTGGTTTGCCTTAAGAGAAATGCAATCGGCATGCAATAGACA 2794
QY 2169 TAGCGGTGATGGCCCTGAAGATGATGGAGTCTC-----TCTGATGAAGTTATGCTCTCCCC 2222
Db 2795 TTGCCAAGATGGCTTGGAAATCTCAGCTTCATGGGACCTTTGAGCTGGAGCATCTTC 2854
QY 2223 ATGAGAACCTATCAAGATGCGAAATTTGGACTGCACTCTGGATCAGTTTTCCTGGCGTCG 2282
Db 2855 CTGGCTCCCAATATGATTCGCATTTGGAGTTTCACTCTGGTCCCTGCTGCTGGAGTTG 2914
QY 2283 TTGGAGTTAAATGCCCCGTTACTGCTTTTTCGAAACAATGTCACTCTGGCTGCTTACAAAT 2342
Db 2915 TGGGAATCAAGATGCCCTCGTTATTGCTATTGAGGATACGGTCAACACAGCCCTCTAGGA 2974
|||||

	Score 21.4	Pos	Best Local Similarity 49.98	Pred. No. 9.2e-15	Mismatches 231	Indels 0	Gaps 0
Matches 230	Conservative	0	Mismatches 231	Indels 0	Gaps 0		
QY	1945	CAAGAAATTCAGTAATGTCACCATGCTCTTCTCAGACATCGTGTGGGTTCACTGCGCATCTG	2004				
Db	1524	CGAGAAACATGACCAACGTGAGCATCTGTTTGCTGACATCGAGGCTTCACACGCTGGC	1583				
QY	2005	CTCCAGTGGTCAACGGTCGAGGTCAACCATGCTCAATGCATGTACACTGCTTTCGA	2064				
Db	1584	GTCCCAATGTCATGTCACAGGAACGTGTCATGACCCCTCAACGAGCTCTTCGCCCGCTTTGA	1643				
QY	2065	CGACGAGTGTGGAGAGCTGGATCTCTACAGGTGGAGACCATTGGCGATGCCATTTCGT	2124				
Db	1644	CAAGCTGGCGCGAGAAATCACCTGTGTACCTAATAAGATCCTTGGGATTGTTATTACTG	1703				
QY	2125	AGCTGGGGGATTACAAAGAGAGTATCATCTGCTGTTCAGATACGCTGATGGCCCT	2184				
Db	1704	CGTCTCGGGCTGCCGTGAACCAAGGGCTGACACGCCCATCTGCTGTGGAGATGGCAT	1763				
QY	2185	GAAGATGATGGAGCTCTCTGATGAAGTTATGTCTCCCATGGAGAACCTATCAAGATGCG	2244				
Db	1764	GGACATGATCGAGGCCATCTCTGTGTGTCGGGAGGTGACAGGGGTGAACCTGAACATGCG	1823				
QY	2245	AATTGGACTGCACCTCTGGATCAGTTTTTGCTGCGCTGGTGGAGTTAAAATGCCCGTTA	2304				
Db	1824	TGTGGGAATTCACAGCGGGGAGTACACTCGGTGTCTTTGGTCTCAGGAAGTGGCAGTT	1883				
QY	2305	CTGCTCTTTTGGAAAAAATCTCACTCTGGCTTAACAAATTTGAGTCTCTGCAGTGTACCACG	2364				
Db	1884	CGAGCTCTGTCTACAGATGTCACGCTAGCCACCACATCGAGGCTGCGCGCAAGGCAGG	1943				
QY	2365	AAAAATCAATGTGAGCCCAACAACCTTACAGATTACTCAAG	2405				
Db	1944	ACGATCCACATCACCAGGCTACACTCAACTACCTGAAATG	1984				

Result No.	Score	Query Match	Length	DB	ID	Description
1	2443	100.0	2443	21	AAZ88939	Human soluble guan
c	2423	99.2	3196	22	AAH93832	Human EST-derived
3	570	23.3	1302	22	AAC33895	Human colon cancer
4	434.8	17.8	459	21	AAH00760	Human secreted pro
5	371.4	15.2	2850	23	ABL02685	Drosophila melanog
6	200.2	8.2	2954	24	AAS34827	Human DNA sequence
7	192.6	7.9	2715	21	AAZ51684	Human cyclic nucle
8	192.6	7.9	3015	21	AAZ88938	Human soluble guan
9	191	7.8	4742	23	ABV21348	Human prostate exp

Qy	730	CAAAAGCTTTTCCCTTTTCATATAATATTGACCGGGACCTAGTGGTCACCTCAGTGTGGCAA	789
Db	2423	CAAAAGCTTTTCCCTTTTCATATAATATTGACCGGGACCTAGTGGTCACCTCAGTGTGGCAA	2364
Qy	790	TGCTATATACAGAGTCTCCCCAGCTCCAGCCTGGGAATTGCGCCTTCTGTCGTCTT	849
Db	2363	TGCTATATACAGAGTCTCCCCAGCTCCAGCCTGGGAATTGCGCCTTCTGTCGTCTT	2304
Qy	850	CTCGCTGGTTCGTCTCATATTGATATTAGTATTTCCATGGGATCCTTTCTCACATCAATAC	909
Db	2303	CTCGCTGGTTCGTCTCATATTGATATTAGTATTTCCATGGGATCCTTTCTCACATCAATAC	2244
Qy	910	TGTTTTTGTATTGAGAAGCAAGGAAGATTGTTGGATGTGGAGAAATTAGAATGTGAGGA	969
Db	2243	TGTTTTTGTATTGAGAAGCAAGGAAGATTGTTGGATGTGGAGAAATTAGAATGTGAGGA	2184
Qy	970	TGAACTGACTGGGACTGGAGATCAGCTAGCTGTTAC--GTCTCAAGGGTCAAAATGATCTACTTAC	1028
Db	2183	TGAACTGACTGGGACTGGAGATCAGCTAGCTGTTACAGTCTCAAGGGTCAAAATGATCTACTTAC	2124
Qy	1029	CTGAAGCAGATAGCATACTTTTCTATGTTTCACCAAGTGTCAATGAACTTGGACGATTTGA	1088
Db	2123	CTGAAGCAGATAGCATACTTTTCTATGTTTCACCAAGTGTCAATGAACTTGGACGATTTGA	2064
Qy	1089	CAAGGAGAGGGCTGTATCTAAAGTCACATCCCTCTGCGATGTCGACGCGGATCTTGTTC	1148
Db	2063	CAAGGAGAGGGCTGTATCTAAAGTCACATCCCTCTGCGATGTCGACGCGGATCTTGTTC	2004
Qy	1149	TTTTTGGAGAAACAATTTAGAGAGGAATACAAACTCACCCAAAGAACTTGGAAATCCTCACTG	1208
Db	2003	TTTTTGGAGAAACAATTTAGAGAGGAATACAAACTCACCCAAAGAACTTGGAAATCCTCACTG	1944
Qy	1209	ACAGGCTACAGCTCACGCTTAAGAGCCCTGGAAGATGAAGAAGAAAGACAGACACATTGC	1268
Db	1943	ACAGGCTACAGCTCACGCTTAAGAGCCCTGGAAGATGAAGAAGAAAGACAGACACATTGC	1884
Qy	1269	TGTATCTGCTCTTCCCTCGCTGTTGCCAATGAGCTGCGGCAACAGCGTCCAGTGCCTG	1328
Db	1883	TGTATCTGCTCTTCCCTCGCTGTTGCCAATGAGCTGCGGCAACAGCGTCCAGTGCCTG	1824
Qy	1329	CCAAAGATATGACAATGTGACCATCCTCTTTAGTGGCATTTGGGCTTCAATGCTTTCT	1388
Db	1823	CCAAAGATATGACAATGTGACCATCCTCTTTAGTGGCATTTGGGCTTCAATGCTTTCT	1764
Qy	1389	GTAGCAAGCATGCATCTGGAGAAGGCCATGAAGATCGTCAACCTCCTCAACGACCTCT	1448
Db	1763	GTAGCAAGCATGCATCTGGAGAAGGCCATGAAGATCGTCAACCTCCTCAACGACCTCT	1704
Qy	1449	ACACCAAGATTGACACACTGACTGATTTCCCGGAAAAACCCATTTGTTTAAAGGTGGAGA	1508
Db	1703	ACACCAAGATTGACACACTGACTGATTTCCCGGAAAAACCCATTTGTTTAAAGGTGGAGA	1644
Qy	1509	CTGTTGGTGACAAGTATATGACAGTGAAGTGGTTTACCAGAGCCATGCATTCACCATGCAC	1568
Db	1643	CTGTTGGTGACAAGTATATGACAGTGAAGTGGTTTACCAGAGCCATGCATTCACCATGCAC	1584
Qy	1569	GATCCATCTGCCACCTGGCCCTCGACATGATGAAATTCGTGCGCAGGTTCAAGTAGATG	1628
Db	1583	GATCCATCTGCCACCTGGCCCTCGACATGATGAAATTCGTGCGCAGGTTCAAGTAGATG	1524
Qy	1629	GTGAATCTGTTTCAGATAACAATAGGGATACACACTGGAGAGTAGTATACAGGTGTCATAG	1688
Db	1523	GTGAATCTGTTTCAGATAACAATAGGGATACACACTGGAGAGTAGTATACAGGTGTCATAG	1464
Qy	1689	GACAGCGGATGCCCTCGATCTGCTTTTGGGGAATACTGTCAACCTTCACAAAGCCGAACAG	1748
Db	1463	GACAGCGGATGCCCTCGATCTGCTTTTGGGGAATACTGTCAACCTTCACAAAGCCGAACAG	1404
Qy	1749	AAACCCACAGAGAAAAGGGAAAAATAAATGTGTCGAATATACATACAGATGCTTATGT	1808
Db	1403	AAACCCACAGAGAAAAGGGAAAAATAAATGTGTCGAATATACATACAGATGCTTATGT	1344
Qy	1809	CTCCAGAAAATTCAGATCCAAATTCACCTTGGAGCACAGAGGCCCGCATCGCAAGG	1868

QY 250 TCGAAGCAAGTCCCTCAATCTCAATGCTGGAGAAATCCCTCCAAATGTTGGGAAGATGT 309
 |||||
 Db 264 TCGAAGCAAGTCCCTCAATCTCAATGCTGGAGAAATCCCTCCAAATGTTGGGAAGATGT 323
 |||||
 QY 310 TTTCTGCTTTTGGCCAGAAATCTGGTTATCATACAAATCTTGGCTGGCTCTAATCT 369
 |||||
 Db 324 TTTCTGCTTTTGGCCAGAAATCTGGTTATCATACAAATCTTGGCTGGCTCTAATCT 383
 |||||
 QY 370 CAGAGAAATTTACAGAACTTGTATGCTCTGCAGCAGCACCTTGTCTACCATCTACCCAGG 429
 |||||
 Db 384 CAGAGAAATTTACAGAACTTGTATGCTCTGCAGCAGCACCTTGTCTACCATCTACCCAGG 443
 |||||
 QY 430 AATGCGTGCACCTTCC 445
 |||||
 Db 444 AATGCGTGCACCTTCC 459

RESULT 5

ABL02665
 ID ABL02665 standard; cDNA; 2850 BP.

XX
 AC ABL02665;
 XX

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 2477.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB58562.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Claim 1; SEQ ID NO 2477; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2850 BP; 734 A; 756 C; 788 G; 572 T; 0 other;

Query Match

Best Local Similarity 15.2%; Score 371.4; DB 23; Length 2850;

Matches 694; Conservative 58.1%; Pred. No. 4.4e-86;

Mismatches 0; Mismatches 471; Indels 30; Gaps 1;

QY 705 CACGCAATCAGCCCATATACATCTCTGCAAAAGCTTTTCTTTTCATATATATTTGACCCGG 764
 |||||
 Db 1276 CACTCATATCGCCGCCACTTCTTCAAGGTGTTCCCTTCCACCTGATGTTTCGATAGGC 1335
 |||||
 QY 765 ACCTAGTGTCACTCAGTGTGGCAATGCTATATACAGAGTTCTCCGCCAGCTCCACCCCTG 824
 |||||
 Db 1336 AATGAAATCTGTTCAGCGCCGCAAGGCTGCTCTCGAGTAATTTCCAGAGTTGCCGAGG 1395
 |||||
 QY 825 GGAATTTGACGCTTCTGCTGCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 884
 |||||
 Db 1396 AGACTGTTCCCTTAATAGAGTGGTGGAGCCATCCGCCGCCACCTGCGAGCTCAACTTCG 1455
 |||||
 QY 885 ATGGGATCCTTTCTCACATCAATACATCTTTTGTATTTGAGAAGCAAGAGGATTTGTTG 944
 |||||
 Db 1456 AGAACATCTGTTCCCATATAAACACCATCTATGCTCTCGAGACACCTCAGGGGGCCATGA 1515
 |||||
 QY 945 ATGTGGAGAAATAGNAATGTGAGGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
 |||||
 Db 1516 GC-----AGCCGGCAGCAGCAGCGCTTTTCTGAGAC 1545
 |||||
 QY 1005 TCAAGGGTCAAAATGATCTACTTACCTGAAGCAGATAGCATACTTTTCTATGTTTACCACAA 1064
 |||||
 Db 1546 TGAAGGGCAATGATGTACATTCGGAAACGGATCGGATCTGTTCAATGCTATCCCA 1605
 |||||
 QY 1065 GTGTCATGAACCTGGACGATTTGACAAAGGAGAGGCTGTATCTAAGTGCATCCCTCTGC 1124
 |||||
 Db 1606 GTGTCATGAATCTGGATGATCTGACCAAGAAGGACTCTACATCTCGGATGTGCCCTGC 1665
 |||||
 QY 1125 ATGATGCCACGCGCATCTGTTCTTTTGGGAGAACATTTAGAGAGGATACAACTCA 1184
 |||||
 Db 1666 ACAGCGTCCAGGATTTGGTCTCTCCGAGAAAGTTCGAGGGCGGAGTACAAGCTCA 1725
 |||||
 QY 1185 CCAAGAACTGGAAATCTCACTGACGCTCAGCTCAGCTTAAGAGCCCTGGAAAGATG 1244
 |||||
 Db 1726 CCAAGAACTGGAAATGCTCAGCAGACATTCGCGATCTGGAGAGC 1785
 |||||
 QY 1245 AAAAGAAAAGACAGACATTTGCTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1304
 |||||
 Db 1786 AGAAGCAGAAGACCGACAGCTGCTCTATTCGGTGTCTGCAAGTGGTGGCAATGAGT 1845
 |||||
 QY 1305 TCGGCGACAAGCGTCCAGTCCGCTGCCAAAGATATGACAATGTGACCATCCCTTTAGTG 1364
 |||||
 Db 1846 TAGGACATCAGCGTCTCTGTGCGCCCAACCGTACGCTCCGAGCTTGTGTTCTCCG 1905
 |||||
 QY 1365 GCATTGTGGGCTTCAATGCTTCTGTAGCAAGCATGCTCTGGAGAGGAGGCGCATGAAGA 1424
 |||||
 Db 1906 GCATGTTGGGCTTGGCCAAATCTGTGCGGCCCAACACAGATCCCGAGCGGCCCATGAAA 1965
 |||||
 QY 1425 TCGTCAACCTCTCAACGACCTCTACACCAAGATTTGACACACTGACTGATTCGCCGAAA 1484
 |||||
 Db 1966 TCGTGAATAATGCTGAATGAATCTACACGCTCTCGATGCTCTGACCGACTCCAAGCGAA 2025
 |||||
 QY 1485 ACCCATTTTATAGGTGGAGACTGTTGGTGACAAGTATATGACAGTGTGTTTAC 1544
 |||||
 Db 2026 ACCTGAACGTGTACAGGTGTTGAACAGTTGGCGATAAGTACATGGCCGTTTCCGGACTGC 2085
 |||||
 QY 1545 CAGAGCCATATTCACCATGCACCATCCATCTGCCACCTGCCCTTGGCATGATGGAAA 1604
 |||||
 Db 2086 CCGATCACTCGCAGGATCATGCCAAGTGCATGCGCAGAGTGGCCCTCGATATGATGACA 2145
 |||||
 QY 1605 TTGCTGGCCAGGTTCAAGTAGATGGTGAATCTGTTTCAGATACATATAGGATACACATG 1664
 |||||
 Db 2146 TGGCCCAAGAACCTCAAAATGGGATCCAAATCCAGTGCAAATCACCATTAGCATTCG 2205
 |||||
 QY 1665 GAGAGGTAGTTACAGGTGTCATAGACAGCGGATGCTCGATGCTCTCTTTTGGGAATA 1724
 |||||
 Db 2206 GCGAAGTGTGACTGAGTGTGCGAACCCGAGTCCACGCTACTGTCTGTTCGGAATA 2265
 |||||
 QY 1725 CTGTCAACCTCAAGCGAACAGACAAACACAGGAGAAAGGAAAAATAAATGTCTCTG 1784
 |||||
 Db 2266 CTGTCAATCTACAGCGGACGAGACCACTGGCGTTCCGGGGCGGCATCAACGTGAGC 2325
 |||||
 QY 1785 AATATACATACAGATGCTTATGTCTCCAGAAAAATTCAGATCCAAATTCCTACTTGGAGC 1844

```

Db      2326 AGAAGACCTATCGGCTCTCTGTATGGCGATCAACAGGATGATTCGTTCCACTTGGAGT 2385
QY      1845 ACAGAGGCCAGCTGCCATGAAGGCAAAAGAAACCAATGCAAGTTTGGTTCT 1899
Db      2386 ACCGGGACCCGTTATCATGAAGGGCAACCCGACGCCCATGCTGGTTCT 2440

RESULT 6
ID      AAS94827
XX      AAS94827 standard; DNA; 2954 BP.
AC      AAS94827;
XX      AAS94827;
DT      14-FEB-2002 (first entry)
XX      Human DNA sequence #82 expressed during foam cell differentiation.
XX      Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW      cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX      Homo sapiens.
XX      OS
XX      PN      WO200177389-A2.
XX      PD      18-OCT-2001.
XX      PF      04-APR-2001; 2001WO-US111128.
XX      PR      05-APR-2000; 2000US-195106P.
XX      PA      (INCY-) INCYTE GENOMICS INC.
XX      PI      Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
XX      PI      Tai J;
XX      PI      WPI; 2002-010925/01.
XX      PT      Composition useful for diagnosis of conditions, disorders or diseases
XX      PT      associated with atherosclerosis, comprises several polynucleotides that
XX      PT      are differentially expressed in foam cell development.
XX      PS      Claim 1; Page 134-135; 315pp; English.
XX      CC      The present invention relates to the isolation of human polynucleotide
XX      CC      sequences that are differentially expressed during foam cell
XX      CC      differentiation. The polynucleotide sequences of the invention or a
XX      CC      composition comprising these polynucleotides are useful as a high
XX      CC      throughput method for detecting altered expression of one or more
XX      CC      polynucleotides in a sample. The polynucleotides can be used in the
XX      CC      diagnosis of disorders associated with foam cell development such as
XX      CC      atherosclerosis, cerebral stroke, and cardiovascular disorders such as
XX      CC      coronary artery disease. The polynucleotide sequences can also be used
XX      CC      as PCR primers and probes. The polynucleotides of the invention are also
XX      CC      useful in gene therapy. AAS94746-AAS95021 represent the human
XX      CC      polynucleotide sequences of the invention which are differentially
XX      CC      expressed during foam cell differentiation.
XX      SQ      Sequence 2954 BP; 735 A; 779 C; 760 G; 680 T; 0 other;

Query Match      8.2%; Score 200.2; DB 24; Length 2954;
Best Local Similarity 55.8%; Pred. No. 1.7e-41;
Matches 457; Conservative 0; Mismatches 338; Indels 24; Gaps 3;

QY      1005 TCAAGGTCAAATGATCTACTTACCTGAAGCAGATAGTACTTTTCTGTACCAAA 1064
Db      1607 TCAAGGCAAAATGATCCATGTTCAGAAATCAAATTCATTTATTTTGGGCTCTCCAT 1666
QY      1065 GTGTCAATGACCTGGAGGCTTTGACAAAGGAGGAGGCTGTATCTAAGTGACATCCCTCTGC 1124
Db      1667 GTGTGACAAAGTTGGATGAACATCATGGCGGAGGGCTACATCTCTCAGACATCCCTATCC 1726
QY      1125 ATGATGCCACGGCGGATCTTTGTTCTTTTGGGAGAACAAATTTAGAGAGGAATACAACTCA 1184

```

```

Db      1727 ATGATGCCACCGAGATGTCAATTTGGTTGGTGAGCAGCAAAAGGCCCAAGATGGGTTGA 1786
QY      1185 CCCAAGAAGCTGGAATCCCTCACTGACAGGCTTACAGCTCACGTTAAGAGCCCTGGAGATG 1244
Db      1787 AGAAAAGGATGATTAATTAAGGCAACTTTAGAAAGAACTCACAGGCCCTGGAGAAAG 1846
QY      1245 AAAAGAAAAGACAGACACATTTCTGTATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1304
Db      1847 AGAAAAGGACAGAGTGGATCTTCTATATTTCTATTTTCTCTCTCTCTCTCTCTCTCTCT 1906
QY      1305 TCGGGCACAAGCGTCCAGTCCCTGCCCCAAAAGATATGACATGTGACCATCTCTCTTTAGTG 1364
Db      1907 TATGGCAAGGGCAGCAAGTACAGGCCAGAAAGTTTGATGATGTGCACCTGCTCTTTTCAG 1966
QY      1365 GCATTTGTGGCTTCAATGCTTTCTGTAGCAAGCATCATCTGAGGAAGAGGAGCCATCAAGA 1424
Db      1967 ACATTTGTGGCTTCAAGCCCATATGTGCCAGTGTAC-----TCCCATCAAG 2014
QY      1425 TCGTCAACCTCTCAACGACCTCTACACAGATTTTGACACACTGACTGATTTCCCGGAAA 1484
Db      2015 TAATCAGCATGCTGAATGAACGTACACACAGATTTGACCACCACTGTTGGATTTTGGAA-- 2072
QY      1485 ACCATTTGTTTATAGGTGGAGACTGTTGGTGACAAGTATATGACAGTGAAGTGGTTTAC 1544
Db      2073 -----TATTTATAGGTGGAACAATAGGTGATGCTCTACTGTGTTCAGCAGGGCTCC 2125
QY      1545 CAGAGCATGCTTACCATGACGATCATCTGCCACCTGGCTTGGACATGATGGA 1604
Db      2126 ACAGAAAAGCTCTGCCATGTAAACCCATGCTCTGTATGGCTTGAAGATGATGGAAC 2185
QY      1605 TTGCTGGCAGGTTCAAGTA---GATGGTGAATCTGTTTCAGATTAACAANTAGGATACACA 1661
Db      2186 TTTCAGAGAGGTGCTGACACCTGATGGAAGACCGATTTCAGATGAGGATAGGAATCACT 2245
QY      1662 CTGGAGAGTAGTTACAGGTGTCATAGGACAGCGGATGCTTCGATCTCTTTTGGGA 1721
Db      2246 CAGGCTCCCTGCTGGCTGGAGTTGTTGGGGTGGCAATGCGACGTTATTCCTTTTGGAA 2305
QY      1722 ATACTCTCACTCACAAGCCGAACAGAACCAACAGGAGAAAGGAAAATAAATGTGT 1781
Db      2306 ATAAATGTCACATGGCAAGCAATTCGAGTCGGGAAGTCACTCCCTCGGCGCATCAATGTC 2365
QY      1782 CTGAATATACATACAGATGCTTTATGCTTCCAGAAAATT 1820
Db      2366 GCCCAACCACTTACCAATTATTAACAGGAGAAAGATT 2404

RESULT 7
AAZ51684
ID      AAZ51684 standard; cDNA; 2715 BP.
XX      AC      AAZ51684;
XX      DT      04-JUL-2000 (first entry)
XX      DE      Human cyclic nucleotide-associated protein-3 (CNAP-3) cDNA.
XX      KW      Cyclic nucleotide-associated protein-3; CNAP-3; human; cytosolic;
XX      KW      anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;
XX      KW      immunomodulatory; anti-asthmatic; anti-anemic; anti-diabetic; diagnosis;
XX      KW      anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;
XX      KW      anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;
XX      KW      anti-fertility; anti-allergic; vasotropic; immunosuppressive;
XX      KW      hypotensive; gene therapy; prevention; treatment; arteriosclerosis;
XX      KW      cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;
XX      KW      neurological; vision; reproductive; smooth muscle; ss.
XX      OS      Homo sapiens.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      440..2512
XX      FT      /*tag= a

```

FT /product= "Human CNAP-3 protein"
 FT /note= "Shares 89% identity to human soluble guanylate
 FT cyclase large subunit"
 FT misc_binding 863..892
 FT /*tag= b
 FT /bound_moiety= "Primer or Probe"
 FT /note= "Useful for amplification or hybridisation
 FT techniques"
 FT
 XX W0200014248-A1.
 XX
 XX 16-MAR-2000.
 XX
 XX 03-SEP-1999; 99WO-US20287.
 XX
 XX 04-SEP-1998; 98US-0148904.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;
 XX
 XX WPI; 2000-256994/22.
 XX P-PSDB; AAY70475.
 XX
 XX Isolated cyclic nucleotide associated proteins useful for preventing,
 XX diagnosing and treating cell proliferative, autoimmune/inflammatory,
 XX neurological, vision, reproductive and smooth muscle disorders -
 XX
 XX Example 3; Page 71-72; 78pp; English.
 XX
 XX The present sequence is the cDNA encoding human cyclic nucleotide
 XX associated protein-3 (CNAP-3), identified in incyte clone 159278,
 XX that is isolated from ADEIN901 cDNA library. It is expressed in
 XX nervous, reproductive, cardiovascular and developmental tissues.
 XX CNAP sequences may be used for prevention, treatment and diagnosis of
 XX diseases associated with altered CNAP expression such as, cell
 XX proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia,
 XX lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/
 XX inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple
 XX sclerosis and psoriasis), neurological disorders (e.g. epilepsy,
 XX Alzheimer's/Parkinson's disease and strokes), vision disorders (e.g.
 XX conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),
 XX reproductive disorders (e.g. infertility, uterine fibroids, ectopic
 XX pregnancies and impotence) and smooth muscle disorders (e.g. angina,
 XX anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also
 XX be used for gene therapy.
 XX
 XX Sequence 2715 BP; 799 A; 604 C; 662 G; 650 T; 0 other;
 XX
 XX Query Match 7.98; Score 192.6; DB 21; Length 2715;
 XX Best Local Similarity 55.58; Pred. No. 1.6e-39;
 XX Matches 447; Conservative 0; Mismatches 334; Indels 24; Gaps 3;
 XX
 XX 1004 CTCAGGGTCAAAATGATCTACTACCTGAAGCAGATAGCATCTTTCTATGTTACCA 1063
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1535 CTCAGGCCCAATGATCTACATGTTGATCCAGTCGCAATCTGTTTGGGGTACCC 1594
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1064 AGTGTCATGAACCTGGACGATTTGACAGAGGAGGGCTGTATCTAAAGTGACATCCCTCTG 1123
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1595 TGTGTGGACAGATTAGAAGATTTTACAGGACGAGGGCTCTACCTCTCAGACATCCCAATT 1654
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1124 CATGATGCCACGGCGGCTCTGTTCTTTTGGGAGACAAATTTAGAGGAGGAATACAACTC 1183
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1655 CACAATGCTAGGAGTGTGGTCTTAATAGGGGAACAGCCCGAGCTCAAGATGGCCTG 1714
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1184 ACCCAAGATCGAAATCCCTCACTGCAGGCTACAGCTACGTTAAGAGCCCTGGAGAT 1243
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1715 AAGAGAGGCTGGGAGCTGAGAGGCTACCTCTGAGCAACCCACCAAGCCCTGGAGGAG 1774
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1244 GAAAGAAAAGACAGACATTCGTGTATTCCTTCCTTCCTCGTGTGTGCAATGAG 1303
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 XX 1775 GAGAAGAAAAGACAGTAGACCTTCTGTGCTCCATATTTTCCCTGTGAGGTTGCTCAGCAG 1834
 XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1304 CTGGGCAAGCGTCCAGTGCCTGCCAAAGATATGACAATGTGACCATCTCTTTAGT 1363
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1835 CTGTGGCAAGGCAAGTTGTGCAAGCCAGAGATTCAGTAATGTACCATCTCTCTCA 1894
 QY 1364 GGCATTGTGGCTTCAATGCTTTCTGTAGCAAGCATGCTCGAGAGGAGCCATGAAG 1423
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1895 GACATGTTGGGTTCATGCTCCAGTGTCTGCTCCAGTGTCTC-----ACGCTCAG 1942
 QY 1424 ATGCTCAACCTCCTCAACGACCTTCAACACAGATTTGACACACTGACTGATTCCTCCGAAA 1483
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1943 GTCATCACCATGCTCAATGCATCTGACATCGCTTCGACCAGCAGTGTGGAGAGCTGA- 2001
 QY 1484 AACCATTTCTTATAGGTGGAGACTGTGTTGGTGACAAGTATATGACAGTGAGTGTTTA 1543
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2002 -----TGTCTACAGGTGGAGACCATTTGGGATGCTTATTGTGTAGCTGGGGATTA 2053
 QY 1544 CCAGAGCCATGCTTACCATGCACGATCCATCTGCCACCTGCGCTTGGACATGATGAA 1603
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2054 CACAAGAGAGTGTACTCATGCTGTTCAGATAGCGCTGATGGCCCTGAAGATGATGGAG 2113
 QY 1604 ATTGCTGGCCAGGTT---CAAGTAGATGGTGAATCTGTTCCAGATAACAATAGGATACAC 1660
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2114 CTCTGTGATGAAGTTATGCTCTCCCATGGAGAACCTATCAAGATGCGAATGGAGTGCAC 2173
 QY 1661 ACTGGAGAGTGTACAGTGTCTATAGGACAGCGGATGCTCGATCTCTCTTTTGGG 1720
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2174 TCTGGATCAGTTTTTGTGCTGCGCTGTTGGAGTTAAATGCGCCGTTACTGTCTTTTGA 2233
 QY 1721 AATACTGTCAACCTCAAGCCGAGAACACAGAGGAGAGGAGGAGGAGGAGGAGGAGGAG 1780
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2234 AACATGCTACTCTGGCTAACAAATTTGAGTCTGCTGAGTACCACGAGGAGGAGGAGGAG 2293
 QY 1781 TCTGAATATATACATACAGATGTCTTA 1805
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2294 AGCCCAACAACCTTACAGATTACTCA 2318
 RESULT 8
 AAZ88938
 ID AAZ88938 standard; DNA: 3015 BP.
 AC AAZ88938;
 XX
 DT 26-MAY-2000 (first entry)
 XX
 DE Human soluble guanylylcyclase alphas DNA.
 XX
 KW Human; guanylylcyclase alphas; hscGcalphas; soluble;
 KW guanylylcyclase beta; antiarteriosclerotic; vasotropic; hypotensive;
 KW gene therapy; arteriosclerosis; restenosis; ischemia; diagnosis;
 -KW peripheral arterial occlusive disease; arterial hypertension; ds.
 XX
 OS Homo sapiens.
 XX
 PN DE19837015-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 14-AUG-1998; 98DE-1037015.
 XX
 PR 14-AUG-1998; 98DE-1037015.
 XX
 PA (VASO-) VASOPHARM BIOTECH GMBH & CO KG.
 XX
 PI Schmidt H, Zabel U, Poller W;
 XX
 DR WPI; 2000-184044/17.
 DR
 DR P-PSDB; AAY51607.
 XX
 XX New human soluble guanylate cyclase alphas/betas and the nucleic acid
 PT encoding the subunits, useful for producing diagnostic antibodies, and
 PT for somatic gene therapy of arteriosclerosis -
 XX

PS Example 1; Page 12-13; 44pp; German.

This invention describes novel purified human soluble guanylate cyclase alpha1/beta1 (hsGCalpha1/beta1). The products of the invention have antiarteriosclerotic, vasotropic and hypotensive activity. Nucleic acid sequences encoding the alpha1 and/or beta1 subunit are useful for somatic gene therapy of arteriosclerosis and restenosis, ischemia (infarct), peripheral arterial occlusive disease and arterial hypertension. Antibodies to hsGCalpha1/beta1 can be used for diagnosis of aberrant hsGCalpha1/beta1 expression in human tissues. This sequence encodes the human soluble guanylylcyclase alpha1 subunit described in the method of the invention.

SQ Sequence 3015 BP; 877 A; 680 C; 716 G; 742 T; 0 other;

Query Match	7.9%;	Score 192.6;	DB 21;	Length 3015;
Best Local Similarity	55.5%;	Pred. No. 1.7e-39;		
Matches 447;	Conservative	0;	Mismatches 334;	Indels 24;
Gaps	3;			

1004	CTCAAGGGTCAAATGATCTACTTTACCTGAAGCAGATAGCATACTTTTTTCTATTGTTCACCA	1061
QY		
Db	1619 CTCAAAGGCCAATGATCTACATTTGTGAATCCAGTGCAATCTGTGTTTTGGSGTCAACC	1678
QY	1064 AGTGTCATGAACCTCGACGATTTGACAAGAGAGGCGCTGTATCTAAGTAGACATGCCTCTG	1123
Db	1679 TGTGTGGACAGATTAGAAGATTTTACAGGACGAGGCGCTCTACCTCTCAGACATCCCAATT	1738
QY	1124 CATGATGCCACGGCCGATCTGTPTCTTTTGGGAGAACAATTTAGAGAGGGAATACAACATC	1183
Db	1739 CACAATGCACTGAGGAGTGTGGTCTTAATAGGGGAACAAGCCGAGCTCAAGATGCGCTT	1798
QY	1184 ACCCAAGAAGCTGAAAATCCTCACTGACAGGCTACAGCTCACGTTAAGAGGCCCTGGAAGAT	1243
Db	1799 AAGAAGAGGCTGGGGAAGCTGAAGGCTACCTTGAGCAAGCCCAAGCCCTGGAGGAG	1858
QY	1244 GAAAGAAAAAGACAGACATTTGCTGTATTTCTTCTCTCTCCGCTGTGTGGCAATGAG	1303
Db	1859 GAGAAGAAAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAG	1918
QY	1304 CTGCGGCACAGGGTCCAGTCGCTGCCAAAGATATGACAATGTAGACCATCTCTTTAGT	1363
Db	1919 CTGTGCAAGGGCAAGTTTGTCAAGGCAAGAAGTTTCAGTAATGTCAACCATGCTCTTCTCA	1978
QY	1364 GGCAATGTGGGCTTCAATGCTTCTCTAGCAAGCATGCACTCGGAGAAGGAGCCATGAAG	1423
Db	1979 GACATCGTTGGGTTCACTGCCATCTGCTCCCAGTGCTC-----ACCGCTGCAG	2026
QY	1424 ATTCGTCAAACCTCCTCAACGACCTCTACCCAGATTTTGACACACTGACTGATTCGCCGAAA	1483
Db	2027 GTCTATCACCATGCTCAATGCACHTGTACATCTCGCTTCGACCAGCAGTGTGGAGAGCTGGA	- 2085
QY	1484 AACCCATTTGTTTATAAGSTGGAGACTGTTGGTGCACAAGTATATGACAGTGAAGTGGTTTA	1543
Db	2086 -----TGCTACAAAGGTGGAGACCATTTGGCGATGCCCTATTTGTAGCTGGGGGATTA	2137
QY	1544 CCNAGCCCATGCAATCACCATGCACGATCCATCTGCCACTTGCCCTTGGACATGTAGGAA	1603
Db	2138 CACAAAGAGAGTGATACTCATGCTGTTTCAGATAGCGCTGATGGCCCTGAAGATGATGGAG	2197
QY	1604 ATTGCTTGGCCAGGTT---CNAAGTAGATGGTGAATCTGTTTCAGATACATAAGGATACAC	1660
Db	2198 CTCCTGTAGTGAAGTTATGCTCCECATGGAGAACCCTATCAAGATGCGAATTTGGCATGCAC	2257
QY	1661 ACTGGAGAGGTAGTTACAGGTGTCAATAGGNACGGGATGCCCTGATACATGCTCTTTTGGG	1720
Db	2258 TCTGGATCAGTTTTTGTGGCGTGTGGAGTTTAAATGCCCCGTGTACTGTCTCTTTTGGGA	2317
QY	1721 AATPACTGTCAACCTCAAGCCCAACAGAACCCACAGGAGAAAAAGGAAAAATAAATGTG	1780
Db	2318 AACAAATGTCACTCTGGCTAACAAATTTGACTCTCGAGTGTACCAAGAAAAATCAATGTC	2377
QY	1781 TCTGAATATACATACAGATGCTCTTA	1805

```
QY 1124 CATGATGCCACGCCGACCTCTGTTCTTTGGGAGAACAAATTTAGAGAGGAATACAAATC 1183
Db 1738 CAAATGCTAGGAGTGTGCTTTAATAGGGAACAAGCCGAGCTCAAGATGCGCTG 1797
QY 1184 ACCCAAGAACTGGAATCTCAGTACAGCTACAGCTCAGCTTAAGAGCCCTGGAGAT 1243
Db 1798 AAGAAGAGCTGGGAGCTGAAGCTACCTTTGAGCAAGCCCAAGCCCTGGAGGAG 1857
QY 1244 GAAAGAAAAGACAGACACATTCGTGTATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 1303
Db 1858 GAGAAGAAAAGACAGACACATTCGTGTATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 1917
QY 1304 CTGCGGCAAGAGCTCCAGTCCGTCGCAAAAGATATGACAATGTGACCATCTCTTTAGT 1363
Db 1918 CTGTGGCAAGGCAAGTGTGCAAGCCAAAGATTCAGTAATTCACCATCTCTCTCTCTCA 1977
QY 1364 GGCATTGTGGCTCAATCTTCTCTAGCAAGCATGCTCTGAGAGGAGGAGCATGAAG 1423
Db 1978 GACATGTTGGTTCACCTGCCATCTGCTCCAGTCTC-----ACCGCTGCAG 2025
QY 1424 ATCGTCAACCTCCTCAACGACCTCTACACAGATTTGACACACTGACTGATTCCTCGGAAA 1483
Db 2026 GTCATCACCATGCTCAATGCACCTGTACACTCGCTTCGACACGAGTGGAGAGCTGGA- 2084
QY 1484 AACCAATTTGTTAAGTGGAGACTGTTGGTGACAAGTATATGACAGTGTGAGTGTGTTA 1543
Db 2085 -----TGCTACAAAGTGGAGACCATTTGGCGATGCTTATGTGTGCTGGGGGATTA 2136
QY 1544 CCAGAGCCATGCTACCATGACGACATCCATCTGCCACCTGGCCCTGGACATGATGAA 1603
Db 2137 CACAAGAGAGTACTCATCTGTTTCAGATAGCCCTGATGGCCCTGAAGATGATGGAG 2196
QY 1604 ATGCTGGCCAGGTT---CAAGTAGATGGTGAATCTGTTCAAGATAACAATAGGATACAC 1660
Db 2197 CTCTCTGATGAAGTTATGCTCCCTCCAGGACCTATCAAGATGCGAATGGAGCTGCAC 2256
QY 1661 ACTGGAGAGTGTACAGGTGTCATAGGACAGCGGATGCTCGATGACTGCTTTTGGG 1720
Db 2257 TCTGGATCATGTTTGTCTGGCGTGTGGAGTTAAATGCCCGCTACTGCTCTCTTGA 2316
QY 1721 ATACTGTCAACCTCAAGCCGACAGAAACACAGAGGAGAAAGGAGAAATAAATATG 1780
Db 2317 ACAAATGTCACCTGCTGCTACAAATTTGAGTCTCTGAGTGATCCAGCAAGAAATCAATGTC 2376
QY 1781 TCTGAATATACATACAGATGCTCTTA 1805
Db 2377 AGCCCAACAACATACAGATTAATCTCA 2401
```

RESULT 10

ABV21746
ID ABV21746 standard; cdna; 4742 BP.

XX AC

XX ABV21746;

DT 13-SEP-2002 (first entry)

XX DE

Human prostate expression marker cdna 21737.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PN 23-APR-2001.

XX PD 20-FEB-2001; 2001WO-0505171.

XX PF 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JB;
XX WPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX Claim 1; Page 3671-3672; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX SQ

Sequence 4742 BP; 1450 A; 933 C; 998 G; 1355 T; 6 other;

Query Watch 7.8%; Score 191; DB 23; Length 4742;

Best Local Similarity 55.4%; Pred. No. 5.5e-339;

Matches 446; Conservative 0; Mismatches 335; Indels 24; Gaps 3;

QY 1004 CTCAGGGTCAATGATCTACTTACCTGAAGCAGATAGATGATCTTTTCTATGTTCCACCA 1063

Db 1618 CTCAGGGTCAATGATCTACTTACCTGAAGCAGATAGATGATCTTTTCTATGTTCCACCA 1677

QY 1064 AGTGTATCACTGAGGAGGCTGTACAGGAGGAGGCTGTATCTAGTACATCTCTCTG 1123

Db 1678 TGTGTGACAGATTAAGAGATTTTACAGGAGGAGGCTCTACCTCTCAGACATCCCAAT 1737

QY 1124 CATGATGCCACGCGCATCTGTTCTTTGGGAGAACAAATTTAGAGAGGAGAAATCAATC 1183

Db 1738 CACATGCACTGAGGATGTTGTTTAAAGGAGAAAGCCGAGCTCAAGATGGCCTG 1797

QY 1184 ACCCAAGAACTGGAATCTCTACTGACAGGCTACAGCTTAAAGAGCCCTGGAGAT 1243

Db 1798 AAGAAGAGGCTGGGAGGCTGAAGGCTACCTTTGAGCAAGCCCAAGCCCTGGAGGAG 1857

QY 1244 GAAAGAAAAGACAGACACATTCGTGTATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 1303

Db 1858 GAGAAGAAAAGACAGACATTCGTGTATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 1917

QY 1304 CTGCGGCAAGGCTCCAGTCCCTGCTGCAAAAGATATGACAATGTGACCATCTCTTTAGT 1363

Db 1918 CTGTGGCAAGGCAAGTGTGCAAGCCAAAGAGTTCAGTAATGTCCACATGCTCTCTCA 1977

QY 1364 GGCATTGTGGCTTCAATCTTCTCTAGCAAGCATGCTCTGAGAGGAGGAGCATGAAG 1423

Db 1978 GACATGTTGGTTCACCTGCCATCTGCTCCAGTCTC-----ACCGCTGCAG 2025

QY 1424 ATCGTCAACCTCCTCAACGACCTCTACACAGATTTGACACACTGACTGATTCCTCGGAAA 1483

Db 2026 GTCATCACCATGCTCAATGCATCTGACTCTGCTCTGAGGAGGAGTGTGAGAGCTGGA- 2084

QY 1484 AACCCATTTGTTTATAAGGTGAGACTGTTGGTGACAAGTATATGACATGATGAGTGTGTTA 1543

Db 2085 -----TGCTACAAAGTGGAGACCATTTGGGATGCTCTTTGTGTAGTGGGGGATTA 2136

(EOSB-) EOS BIOTECHNOLOGY INC.

Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

WPI; 2002-471335/50.

P-PSDB; ABG61878.

Detecting a prostate cancer-associated transcript in a cell in a patient, useful for diagnosing prostate cancer (PC) or screening modulators of PC, by determining if prostate cancer-associated genes are expressed in a prostate tissue

Claim 22; Page 362-363; 436pp; English.

The present invention relates to methods of detecting a prostate cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with prostate cancer-associated polynucleotides (designated PC genes) that selectively hybridize to a sequence that is at least 80% identical to them. The prostate cancer-associated polynucleotide sequences are differentially expressed in prostate tumour tissue or in prostate cancer and are derived from the tissues of various organisms such as humans or other mammals (e.g. mice, sheep and dogs). The methods of the invention are useful for diagnosing and treating prostate cancer in mammals. The prostate cancer-associated genes are useful for diagnosing or treating prostate cancer, as well as for identifying modulators of prostate cancer or agents that inhibit prostate cancer. The nucleic acid sequences are particularly useful in gene therapy, as a vaccine or in antisense applications.

ABK92115-ABK92263 represent prostate cancer-associated polynucleotide sequences.

Sequence 3004 BP; 874 A; 679 C; 710 G; 741 T; 0 other;

Query Match

Best Local Similarity 7.6%; Score 186.2; DB 24; Length 3004;

Matches 443; Conservative 0; Mismatches 338; Indels 24; Gaps 3;

QY 1004 CTCAGGTCGAATGATCTACTTACCTGAAGCAGATGACATACACTTTTCTATGTTTCACCA 1063
DB 1609 CTTAAAGGCCAAATGATCTATGTTTGAATCCAGTGCATATCTTTTGGGGTCAACC 1668
QY 1064 AGTGTGATGAACCTGGAGGCTTGACAGGAGGCTGTATCAAGTGACATCCCTCTG 1123
DB 1669 TGTGTGGACAGATTAGAAGATTTTACAGAGAGGGCTCTACCTCTCAGACATCCCAAT 1728
QY 1124 CATGATGCCACGCGCATCTTTCTTTTGGGAGAACAAATTTAGAGAGGAATACAACTC 1183
DB 1729 CACAATGCACTGAGGATGTGTCTTAATAGGGGAACAAGCCGAGCTCAAGATGGCCTG 1788
QY 1184 ACCCAAGACTGGAAATCTCACTGACAGGCTACAGCTCAGTTAAGAGCCCTGGAAGAT 1243
DB 1789 AAGAAGAGGCTGGGGAGCTGAAGGCTACCCCTTGAGCAAGCCCAAGCCCTGGAGAG 1848
QY 1244 GAAAGAAAAGACAGACACATTTGCTGTAATCTGCTCTCTGCTGCTGCTGCTGCTGCTG 1303
DB 1849 GAGAGAAAAGACAGATGACCTCTGTGCTCTCCTATATTTCCCTGTGAGTGTCTCAGCAG 1908
QY 1304 CTGGGCAACAAGCTCCAGTCCCTGCCAAAGATATGACATGTGACCATCTCTTTAGT 1363
DB 1909 CTGTGGCAAGGCAAGTGTGCAAGCCAAAGATTCAGTAATGTCAACCATGCTCTCTCA 1968
QY 1364 GCAATGTGGCTTCAATGCTTTCTGTGAGCAAGCATGCTGGAAGAGGAGCCATGAAG 1423
DB 1969 GACATCGTGGGTTTCACTGCGCATGCTGCTCCAGTGCTC-----ACCGCTGCAG 2016
QY 1424 ATCGTCACCTCCCTCAAGGCTCTACACAGATTTGACACATGACTGATTCGCGGAA 1483
DB 2017 GTCATCACCATGCTCAATGCACTGTACACTGCTCTCGACAGCATGTGGAGAGTGA- 2075
QY 1484 AACCCATTGTGTTAAGGTGGAGACTGTTGGTGCACAGATATATGACATGAGTGGTTTA 1543
DB 2076 -----TGCTACAAAGTGGAGACCATTCGGATGCCCTATTTGTGTGGCTGGGGGATTA 2127

QY 1544 CCAGAGCCATGCATTCAACCTGACGATCCATCTGCGCACCTGCGCTTGGACATGATGAA 1603
DB 2128 CACAAAGAGAGTGTATCTACTGCTCTTTCAGATAGCGCTGATGGCCCTGAAGATGATGAG 2187
QY 1604 ATTGCTGCCAGGTTCAAGTA--GATGGTGAATCTGTTTCAGATAACAATAGGGATACAC 1660
DB 2188 CTCCTGTATGAAGTATGCTCTCCCATGGAGAACCTATCAAGATGCGAATGGAGCTGCAC 2247
QY 1661 ACTGGAGAGGTAGTTACAGAGTGTCTATAGGACAGCGGATGCGCTCGATACTGTCTTTTGGG 1720
DB 2248 TCTGATCAGTTTTTGTGCTGGCTGCTGTTGAGTAAATGCCCCGTTACTGTCTTTTGA 2307
QY 1721 AATCTGTCAACCTCACAAGCCGGAACACAAACCACAGAGAAAGGGAATAAATGTCG 1780
DB 2308 AACATGTCACTCTGGCTTAACAAATTTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2367
QY 1781 TCTGAATATACATACAGATGCTTTA 1805
DB 2368 AGCCCAACAACCTTACAGATTACTCA 2392

RESULT 14

ABK92259

ID ABK92259 standard; DNA; 3004 BP.

AC ABK92259;

DT 15-AUG-2002 (first entry)

DE Prostate cancer-associated DNA sequence #145.

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;

OS Mammalia.

PW WO200230268-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US32045.

PR 13-OCT-2000; 2000US-0687576.

PR 08-DEC-2000; 2000US-0733288.

PR 24-JAN-2001; 2000US-0733742.

PR 16-MAR-2001; 2001US-263957P.

PR 16-MAR-2001; 2001US-276791P.

PR 06-APR-2001; 2001US-281922P.

PR 30-APR-2001; 2001US-286214P.

PR 04-MAY-2001; 2001US-288589P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

WPI; 2002-471335/50.

P-PSDB; ABG61940.

Detecting a prostate cancer-associated transcript in a cell in a patient, useful for diagnosing prostate cancer (PC) or screening modulators of PC, by determining if prostate cancer-associated genes are expressed in a prostate tissue

Claim 22; Page 421; 436pp; English.

The present invention relates to methods of detecting a prostate cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with prostate cancer-associated polynucleotides (designated PC genes) that selectively hybridize to a sequence that is at least 80% identical

CC	to them. The prostate cancer-associated polynucleotide sequences
CC	are differentially expressed in prostate tumour tissue or in
CC	prostate cancer and are derived from the tissues of various
CC	organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC	The methods of the invention are useful for diagnosing and treating
CC	prostate cancer in mammals. The prostate cancer-associated genes are
CC	useful for diagnosing or treating prostate cancer, as well as for
CC	identifying modulators of prostate cancer or agents that inhibit
CC	prostate cancer. The nucleic acid sequences are particularly useful
CC	in gene therapy, as a vaccine or in antisense applications.
CC	ABK92115-ABK92263 represent prostate cancer-associated polynucleotide
CC	sequences.
XX	
SQ	Sequence 3004 BP; 874 A; 679 C; 710 G; 741 T; 0 other;
	Query Match 7.6%; Score 186.2; DB 24; Length 3004;
	Best Local Similarity 55.08; Pred.No. 7.8e-38;
	Matches 443; Conservative 0; Mismatches 338; Indels 24; Gaps 3;
QY	1004 CTCAAGGGTCAAAATGATCTACTTACCTCGAACAGATAGCATCTTTTTCATGTTCACCA 1063
DB	
	1609 CTCAAAGGCCAAATGATCTACATTGTTGAATCCAGTGCAATCTGTGTTTTGGGGTCAACC 1668
QY	1064 AGTGTGATGAACCTGACAGATTTTGACAGGAGAGGGCTGTATCTTAAGTGCATCCCTCTG 1123
DB	
	1669 TGTTGTGGCAGATTAGAAGATTTTACAGGACGAGGGCTCTACCTCTCAGACATCCCCAATT 1728
QY	1124 CATGATGCCACGGCGCTGTCTTTCTTTTCGGGAGAACAATTTAGAGAGGAATACAAAATC 1183
DB	
	1729 CACANTGCATGAGGATGTGGTCTTATPAGGGGAACACGCCGAGCTCAAGATGGCGCTG 1788
QY	1184 ACCCAAGAACTGGAAATCCTCACTGCAGGCTACAGCTCACGTTAAGAGGCCCTGGAAAGAT 1243
DB	
	1789 AAGAAGAGGCTGGGAAGCTGAAAGCTTACCTTTGAGCAAGCCCACCAAGCCCTGGAGGAG 1848
QY	1244 GAAGAAGAAAGACACAGCATTTGCTGTATTCTGTCTCCTTCCTCGGCTGTTGCCAATGAG 1303
DB	
	1849 GAGAAGAAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAG 1908
QY	1304 CTCGGGCACAAGCGTCCAGTGCCTGCCAAAAGATATGACAACTGACACATCTCTTTTAGT 1363
DB	
	1909 CTGTGGCAAGGGCAAGTTGTGCAAGCCACAGAAGTTCAGTATGTGCACCTAGCTCTTCTCA 1968
QY	1364 GGCAATTGTGGGCTCAATGCTTTCTGTAGCAAGCATGCATCTGAGAAAGAGAGCCATGAAG 1423
DB	
	1969 GACATCGTTGGGTTCACTGCCATCTGCTCCAGTGCTC-----ACCGCTCGAG 2016
QY	1424 ATGCTCAACCTCCTCAACGACCTCTACACAGATTTTGACACATGACTGATATCCCGGAAA 1483
DB	
	2017 GTCATCACCATGCTCAATGCACCTGTACACTCGCTTCGACCAGCAGTGTGGAGAGCTGGA- 2075
QY	1484 AACCCATTTCTTTTAAAGGTGGAGACTGTTGGTGCACAAAGTATATGACAGTCAAGTGGTTTA 1543
DB	
	2076 -----TGTCTACAGGTGGAGACCAATTCGCGATGCTTATTTGTGGGCTTGGGGGATTA 2127
QY	1544 CCAGAGCCATGCATTACCATGCACGATTCATCTGCCACCTGGCCTGGACATGATGAA 1603
DB	
	2128 CACAAAGAGAGTATACTCATGTGTTTCATATAGCGCTGATGGCCCTGAAGATGATGCGAG 2187
QY	1604 ATTGCTGGCCAGGTTCCAAGTA---GATGGTGAATCTGTTCCAGATAACAATAGGAGTACAC 1660
DB	
	2188 CTCTCTGATGAAGTTATGTCTCCCCATGGAGAACCATTACAAGATGCCAATTTGGAGTCAC 2247
QY	1661 ACTGGAGAGGTAGTTACAGGTGTCAATAGGACACGGGATGCCTCGATCTCTTTTPTGGG 1720
DB	
	2248 TCTGGATCAGTTTTTCTGCGCGTGTGGAGTTTAAATGTCCCGGTTACTCTCTCTTTTGGGA 2307
QY	1721 AATACTGTCAACTCTACAAGCCCAACAGAAACACAGGAGAAAAAGGGAATAAATGATGTG 1780
DB	
	2308 AACAATGTCACTGCGCTACAAAATTTGACTCTCGAGTGTACCAAGAAATCAATGTC 2367
QY	1781 TCTGAATATACATACAGATGTCTTTA 1805

The invention relates to methods for predicting toxic effects of compounds or the progression of these toxic effects by determining the global changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells. Also included are methods of predicting at least one toxic effect of a compound or progression of a toxic effect, preferably the hepatotoxicity of a compound, comprising detecting the level of expression in a tissue or cell sample exposed to the compound of two or more genes listed in the specification, where differential expression of the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilise a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridises to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information, identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell.

CC sample that has been exposed to a compound or agent. Hepatotoxicity
CC is characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent.
XX
SQ Sequence 2335 BP; 660 A; 535 C; 596 G; 544 T; 0 other;

Query Match 6.0%; Score 147.2; DB 24; Length 2335;
Best Local Similarity 51.0%; Pred. No. 9.9e-28;
Matches 517; Conservative 0; Mismatches 443; Indels 54; Gaps 5;

Qy	725	TTCTGCAAGAGCTTTTCTCTTTTCATATATAATTTTGGACGGGACCTAGTGGTCACTCAGTGT	784
Db	771	TTCTGTATGCTTTTCTCTTTCCACATTTGTTGATGAAGCACTAAAGGCTCAAGCAAGCT	830
Qy	785	GGCAATGCTATATACAGAGTTCTCCCCAGCTCCAGCTGGGAACTTGCAGCCCTCTCTCT	844
Db	831	GGAGTGAATATTCAGAAAGTAGTCCCTGGGAATCTTAACCCAGAAAGTTTGCACATAGATGAG	890
Qy	845	GTCTCTTCGCTGGTTCGCTCTCATATATGATATAGTTTCCATGGGATCTTTTCTCACATC	904
Db	891	TATTTTCCATCATCCACCCCTCAAGTTACTTTTCAACATCTCCAGCATCTCGAAGTTCAATT	950
Qy	905	AATACTGTTTTTGTATTGGAAGCAAGGAAGGATTTGTGGATCTGGAGAAATTAGAATGT	964
Db	951	AACAGTCAGTTTGTCTTGAAGACAAGAAAAGAAATGATGCC-----AAAGC	997
Qy	965	GAGGATCAACTGACTGGGACTGAGATCAGCTGCTTAGCTCTCAAGGCTCAAAATGATCTAC	1024
Db	998	AAGGAAGAGCCAGCCGATGCTCAA-----CTCGGGGTCTAGATGATCTGG	1043
Qy	1025	TTACCTGAAGCAGATAGCATACATTTTTTCTATGTTTCCAAAGTGTCAATGAACCTGGAGCAT	1084
Db	1044	ATGAGTCTCTGAGTGCATGATCTCTCATGTTTCCCAAAGTCCGACGCTCGCAAG	1103
Qy	1085	TTGACAGGAGAGGGCTGTATCTAAGTGAACATCCCTCTGCATGATGCCAGCGCGATCTT	1144
Db	1104	CTGGAAGAGAGCAAGATGCATCTTCTGTATCGCTCCGACGACGACCGAGGATCTC	1163
Qy	1145	GTTCCTTTTGGGAAACAATTTAGAGAGGAATACAACTCACCCAAGAACTTGGAAATCCTC	1204
Db	1164	ATCTCTCTCAACCAGCAGAGGCTGGCAGAGATGGAGCTGTCTGCCAACTGGMAAAGAAAG	1223
Qy	1205	ACTGACAGGCTACAGCTCAGTTTAAGAGCCCTTGAAGATGAAAGAAAAGACAGACACA	1264
Db	1224	AAGGAGAGTTGGTGTCTCTTCCAACTCAGCTGGCCATCGAGAAGAAAGACAGAGACC	1283
Qy	1265	TTGCTGTATTCTGTCTCTCCGTCTGTGGCCAACTCAGCTGGGGCAACAGCGTCCAGTG	1324
Db	1284	TTGCTGTATGCCATGTGCTCTGAACATGTGGCCAACTCAGGAGGGCAGAAAGGTG	1343
Qy	1325	CCTGCCAAAAGATATGACAAATGTGACCATCTCTTTAGTGCCATTTGGGGCTTCAATGCT	1384
Db	1344	GCTGCAGGAAATTTGAAACATGTACAACTCTTTTACGCGATGTGTGACATTTACCAAC	1403
Qy	1385	TTCTGTAGCAAGCATGCATCTGGAGAAGGAGCCATGAAGATPGTCAACCTCTCAACGAC	1444
Db	1404	ATCTGTGCAGCCTGTGAACCT-----ATCCAAATCGTGAACATGCTGAATTC	1451
Qy	1445	CTCTACACCAAGATTTGACACACTGACTGATTCGGGAAAACCCATTTGTTTATAGGTG	1504
Db	1452	ATGTACTCCAAGTTTGAACAGGTTAACAGGTGTCCATGA-----TGCTACAAAGTA	1502
Qy	1505	GAGACTCTTGGTGACAAAGTATATACACAGTGAGTGGTTTACCAGAGCCATGCATTCACCAT	1564
Db	1503	GAAACAATAGGGATGCTTACATGGTGGTGGAGTACCAGTACCCTTTGAAAGGCAT	1562
Qy	1565	GCAGATCCATCTGCCACCTGGCCCTTGGACATGA-----TGGAAATTTGCTGGCCAGGTT	1618
Db	1563	GCTCAAGAGTCGCCAATTTTGCTCTGGGATGAGATTTCTGCAAAAAGATGATCAAT	1622
Qy	1619	CAAGTAGATGGTGAATCTGTTTCATATAACAATAGGGATACACACTGGAGAGGTAGTTACA	1678

Accession	Gene	Sequence
1623	Db	CGTGTCACTGGGGAACCTATCCAGATCAGAGTGGGAATCCACACTGGACGACTCTTAGCA 1682
1679	Qy	GGTGTCAATAGGACAGCGGATGCCTCGATACTGTCTTTTGGGAATACTGTCAAC 1732
1683	Db	GGTGTGTGGGAGACAAGATGCCTCGGTACTGCTGTTCGGTGACACTGTAAAC 1736

Search completed: July 1, 2003, 13:26:20
Job time : 357.866 secs


```
QY 1966 GAAGAGGATACAGACTAGTTCAGATTTCTCTTAACAGGTGCCAAGCCAGGAGCAGT 2025
Db 1884 GAAGAGGATACAGACTAGTTCAGATTTCTCTTAACAGGTGCCAAGCCAGGAGCAGT 1943
QY 2026 TCTTCCCTATGGATACAGATTTCTTTGTGCTCTGCTTACATACCCCAAGACTTTCTTCTA 2085
Db 1944 TCTTCCCTATGGATACAGATTTCTTTGTGCTCTGCTTACATACCCCAAGACTTTCTTCTA 2003
QY 2086 GATATATCTCTACTATCCGTTATTAACCTTAGTCTGCTTCTATTAATTTTAGGCT 2145
Db 2004 GATATATCTCTACTATCCGTTATTAACCTTAGTCTGCTTCTATTAATTTTAGGCT 2063
QY 2146 TTAGTATATATCTAAAGTTGGCTTTGATGTGGATGATGAGTTCATGCTGCTTAA 2205
Db 2064 TTAGTATATATCTAAAGTTGGCTTTGATGTGGATGATGAGTTCATGCTGCTTAA 2123
QY 2206 AATCTACTACAAGCATTAACCTAACATGCTGATCTGCAAGTAGTAGGACCAATATAATAT 2265
Db 2124 AATCTACTACAAGCATTAACCTAACATGCTGATCTGCAAGTAGTAGGACCAATATAATAT 2183
QY 2266 TTGTTGAATTTAGTTAAATGAACACTGAACAGTGTGGCCATGCTGATATATTTATATCATG 2325
Db 2184 TTGTTGAATTTAGTTAAATGAACACTGAACAGTGTGGCCATGCTGATATATTTATATCATG 2243
QY 2326 TTTACCAAAATCTGTTAGTTAGTCTCCACATATATGATATATATTTAAATGACTATAATG 2385
Db 2244 TTTACCAAAATCTGTTAGTCTCCACATATATGATATATATTTAAATGACTATAATG 2303
QY 2386 TAATAAAGTTTATATCATGTTGGTGTATATCATTTAGAAATCAATTTCTTAAAGAGT 2443
Db 2304 TAATAAAGTTTATATCATGTTGGTGTATATCATTTAGAAATCAATTTCTTAAAGAGT 2361

RESULT 2
AL548673 919 bp mRNA linear EST 16-FEB-2001
LOCUS AL548673 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1036YN13 5
DEFINITION Prime mRNA sequence.
ACCESSION AL548673
VERSION AL548673.1 GI:12883913
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 919)
AUTHORS Li,W.B., Gruber,C., Jesses,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. 919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1036YN13"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site.1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 252 a 204 c 207 g 254 t
ORIGIN 2 others
```

```
Query Match 37.0%; Score 903.4; DB 9; Length 919;
Best Local Similarity 99.5%; Pred. No. 1e-216;
Matches 915; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 17 CCGCCTCTGCTGGTCCCTTCGCGCGTACCTCTGCGTGGGGGCTGCTCCCGGCTCC 76
Db 1 CCGCCTCTGCTGGTCCCTTCGCGCGTACCTCTGCGTGGGGGCTGCTCCCGGCTCC 60
QY 77 GGTGCAGACACCATGTAGCGGATTTGTGAATCACGCCCTGGAGTTGCTGGTGTATCGGCAAT 136
Db 61 GGTGCAGACACCATGTAGCGGATTTGTGAATCACGCCCTGGAGTTGCTGGTGTATCGGCAAT 120
QY 137 TACGCCCCCGAGGTGTGGGAAGACATCAAAAAGAGGCACAGTTAGATGAGAAGAGCAG 196
Db 121 TACGCCCCCGAGGTGTGGGAAGACATCAAAAAGAGGCACAGTTAGATGAGAAGAGCAG 180
QY 197 TTTCTTGTGAGAATAATATATGATGACTTCCAAAATCTTATGATTTGGTGTGCTGCTCAAGC 256
Db 181 TTTCTTGTGAGAATAATATATGATGACTTCCAAAATCTTATGATTTGGTGTGCTGCTCAAGC 240
QY 257 AAAGTCTCTCAATCTCAATGTGTGGAGAAATCTCCAAAATGTTTGGGAAGATGTTTTCGTC 316
Db 241 AAAGTCTCTCAATCTCAATGTGTGGAGAAATCTCCAAAATGTTTGGGAAGATGTTTTCGTC 300
QY 317 TTTTGGCAAGAAATCTGTTATGATACAACTCTGGTGTCTCTGGCTCTTAATGTCAGAA 376
Db 301 TTTTGGCAAGAAATCTGTTATGATACAACTCTGGTGTCTCTGGCTCTTAATGTCAGAA 360
QY 377 TTTCTACAGAACTTGTATGCTCTGCAGCAGCACTTGTCTACCATCTACCCAGGAATGGT 436
Db 361 TTTCTACAGAACTTGTATGCTCTGCAGCAGCACTTGTCTACCATCTACCCAGGAATGGT 420
QY 437 GCACCTTCTCTTTAGGTGCACCTGATGCAGAAAAGGCAAGGACTCATTTTGCACCTACTAC 496
Db 421 GCACCTTCTCTTTAGGTGCACCTGATGCAGAAAAGGCAAGGACTCATTTTGCACCTACTAC 480
QY 497 TCAGAGAGAGAGAGACTTCAGGATATTTGTCATTTGGAATCATCAAAAGAGTGGCACAACAA 556
Db 481 TCAGAGAGAGAGAGACTTCAGGATATTTGTCATTTGGAATCATCAAAAGAGTGGCACAACAA 540
QY 557 ATCCATGGCACTGAAATAGACATGAAGGTTATTTCAGCAAAAGAAATGAAGAATGTGATCAT 616
Db 541 ATCCATGGCACTGAAATAGACATGAAGGTTATTTCAGCAAAAGAAATGAAGAATGTGATCAT 600
QY 617 ACTCAATTTTAAATGAAGAAAAGAGTCAAAAGAGAGGATTTTATGAAGATCTTGAC 676
Db 601 ACTCAATTTTAAATGAAGAAAAGAGTCAAAAGAGAGGNTTTTATGAAGATCTTGAC 660
QY 677 AGATTTGAAGAAATGTTACCCAGGAATCAGCATCAGCCCATATACATTTCTGCAAGCT 736
Db 661 AGATTTGAAGAAATGTTACCCAGGAATCAGCATCAGCCCATATACATTTCTGCAAGCT 720
QY 737 TTTCTTTTTCATATAATATTTGACCGGACCTAGTGGTCACTAGTGGAATGCTATA 796
Db 721 TTTCTTTTTCATATAATATTTGACCGGACCTAGTGGTCACTAGTGGAATGCTATA 780
QY 797 TACAGAGTTCTCCCCAGCTCCAGCCTGGGAATTCAGCCTTCTGCTGCTTCTTCGCTG 856
Db 781 TACAGAGTTCTCCCCAGCTCCAGCCTGGGAATTCAGCCTTCTGCTGCTTCTTCGCTG 840
QY 857 GTTCGCTCTCATATATGATTTAGTTCATCGGATCCCTTTCTCATCAATACATCTTTTT 916
Db 841 GTTCGCTCTCATATATGATTTAGTTCATCGGATCCCTTTCTCATCAATACATCTTTTT 899
QY 917 GTATTGAGAAGCAAGGAAGG 936
Db 900 GTATTGAGAAGCAAGGAAGG 919

RESULT 3
BQ421149 1073 bp mRNA linear EST 23-MAY-2002
LOCUS BQ421149
DEFINITION AGENCOURT_7766609 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6015435
```



```

5', mRNA sequence.
B0421149
VERSION B0421149.1 GI:21116464
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 1073)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13211 row: 1 column: 04
High quality sequence stop: 617.
Location/Qualifiers
1..1073
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6015435"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 300 a 236 c 262 g 275 t
ORIGIN
Query Match 34.4%; Score 840.6; DB 14; Length 1073;
Best Local Similarity 93.5%; Pred. No. 6.9e-201;
Matches 932; Conservative 0; Mismatches 59; Indels 6; Gaps 5;
QY 559 CCATGGCACTGAATAGACATCAAGGTATTTCAGCAAGAAATGAAGATGTGATCATAC 618
DB 10 CGATGGCACTGAATAGACATCAAGGTATTTCAGCAAGAAATGAAGATGTGATCATAC 69
QY 619 TCAATTTTAAATGAAGAAAGAGTCAAAAGAGGATTTTATGAGATCTTGACAG 678
DB 70 TCAATTTTAAATGAAGAAAGAGTCAAAAGAGGATTTTATGAGATCTTGACAG 129
QY 679 ATTTGAAGAAATGTGTACCCAGCAATCAGCCATACATATACATTTCTGCAAAAGCTTT 738
DB 130 ATTTGAAGAAATGTGTACCCAGCAATCAGCCATACATATACATTTCTGCAAAAGCTTT 189
QY 739 TCCCTTTTCATATAATATTGACCGGGACCTAGTGTCTAGTGTGGAATGCTATATA 798
DB 190 TCCCTTTTCATATAATATTGACCGGGACCTAGTGTCTAGTGTGGAATGCTATATA 249
QY 799 CAGAGTTCTCCCGCCAGCTCCAGCTGGGAATGACGCTTCTGCTCTCTCCGCTGT 858
DB 250 CAGAGTTCTCCCGCCAGCTCCAGCTGGGAATGACGCTTCTGCTCTCTCCGCTGT 309
QY 859 TCGTCTCATATATGATATAGTTTCCATGGGATCCCTTTTCACATCAATACATGTTTGT 918
DB 310 TCGTCTCATATATGATATAGTTTCCATGGGATCCCTTTTCACATCAATACATGTTTGT 369
QY 919 ATTGAGACAGGAGGAGGATTTGTTGGATGTGAGAAATAGATGTGAGGATGACTGAC 978
DB 370 ATTGAGACAGGAGGAGGATTTGTTGGATGTGAGAAATAGATGTGAGGATGACTGAC 429
QY 979 TGGGACTGAGATCAGCTGCTACGCTCAAGGGTCAAAATGATCTACTTACCTGAGGAGAGA 1038

```

```

Db 430 TGGGACTGAGATCAGCTGCTTACGTCTCAAGGGTCAAAATGATCTACTTACCTGAAGCAGA 489
QY 1039 TAGCATACTTTTCTTATGTTTACCAAGTGTCTCAACCTGGAGGATTTTACCAAGGAGAGS 1098
DB 490 TAGCATACTTTTCTTATGTTTACCAAGTGTCTCAACCTGGAGGATTTTACCAAGGAGAGS 549
QY 1099 GCTGTATCTAAGTACATCCCTCTGTCATGATCCGCGCGATCTTGTCTTTTGGGAGA 1158
DB 550 GCTGTATCTAAGTACATCCCTCTGTCATGATCCGCGCGATCTTGTCTTTTGGGAGA 609
QY 1159 ACAATTTAGAGAGGATACAAACTCA-CCCAAGAACTGAAATCTCTACTGACAGGCTAC 1217
DB 610 ACAATTTAGAGAGGATACAAACTCA-CCCAAGAACTGAAATCTCTACTGACAGGCTAC 669
QY 1218 AGCTCAGCTTAAGAGCCCTGGGAAGATGAAAGAAAGACACACATCTGCTGTTATTCG 1277
DB 670 AGCTCAGCTTAAGAGCCCTGGGAAGATGAAAGAAAGACAGCCCTTGTCTGTTATTCG 729
QY 1278 TCCCTTCCCTCCGCTCTGTTGCCAATGAGCTGCGGCACAAAGCGTCCAGTGCCTGCCAAAGAT 1337
DB 730 TCCCTTCCCTCCGCTCTGTTGCCAATGAGCTGCGGCACAAAGCGTCCAGTGCCTGCCAAAGAA 789
QY 1338 ATGACAATGTG-ACCATCTCTTTAGTGGCATTTGGGCTTCAATGCTTTCTGTAGCAGAG 1396
DB 790 GTGACAATGTGAACCATCTCTTAGAGGGCATTTGGGGCTTCAAGGCTTTCTGTACAAAG 849
QY 1397 CATGCATCTGGAGAGGAGGACCATGAAGATCGT-CAACCTCTCTCAAGGACCTCTACACACAG 1455
DB 850 CAGGCATTTGGGAGAGGAGGACCATGAATCTGTCGAACCTCTCTCAAGGAACTCTTACACAG 909
QY 1456 ATTT-GACACATGATGATTTCCCGGAAAAACCCATTTG--TTTATAAGGTGGAGACTGT 1512
DB 910 ATTTGGCAACACCGCATGAATTTCCGGGAGAAAACCTTTTGGTTTATATAGGGGGAAATGCT 969
QY 1513 TGGTGACAAAGTATATGACAGTGTGAGTGTGTTTACCAGAG 1549
DB 970 TGTGCCCCAGTATTTGAACCGGGGTGGTGTATACCCAG 1006

```

```

RESULT 4
B1916583
LOCUS B03178512F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5242779 5',
DEFINITION mRNA sequence.
ACCESSION B1916583
VERSION B1916583.1 GI:16199460
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 809)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11612 row: g column: 04
High quality sequence stop: 804.
Location/Qualifiers
1..809
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5242779"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"

```

FEATURES
source

/note="Organ: brain; Vector: pcMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."

```
BASE COUNT      245 a 165 c 176 g 223 t
ORIGIN
Query Match      32.0%; Score 782.8; DB 13; Length 809;
Best Local Similarity 99.5%; Pred. No. 2.5e-186;
Matches 806; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 1452 CCAGATTGACACACTGACTGATCCCGGAAACCCCAATTTGTTTATAGGTGGAGACTG 1511
DB 1 CCAGATTGACACACTGACTGATCCCGGAAACCCCAATTTGTTTATAGGTGGAGACTG 60
QY 1512 TTGGTGACAAGTATATGACAGTGTGTTTACCAGAGCCCATGCAATCCACCATGCACGAT 1571
DB 61 TTGGTGACAAGTATATGACAGTGTGTTTACCAGAGCCCATGCAATCCACCATGCACGAT 120
QY 1572 CCATCTGCCACTGCCCTTGACATGATGGAATGCTGGCCAGGTTCAGTAGATGTG 1631
DB 121 CCATCTGCCACTGCCCTTGACATGATGGAATGCTGGCCAGGTTCAGTAGATGTG 180
QY 1632 AATCTGTTGACATACATATGAGGATACACACTGGAGAGGTAGTTACAGGTGTCATAGGAC 1691
DB 181 AATCTGTTGACATACATATGAGGATACACACTGGAGAGGTAGTTACAGGTGTCATAGGAC 240
QY 1692 AGCGGATGCCCTGATGCTCTTTTGGGAATGCTCAACTCAAGCCGCAACAGAAA 1751
DB 241 AGCGGATGCCCTGATGCTCTTTTGGGAATGCTCAACTCAAGCCGCAACAGAAA 300
QY 1752 CCACAGGAGAAAGGAAATATATGCTGTGATATACATACAGATGCTTATGTCTC 1811
DB 301 CCACAGGAGAAAGGAAATATATGCTGTGATATACATACAGATGCTTATGTCTC 360
QY 1812 CAGAAATTCAGATCCCAATTCACCTTGGAGCAGAGGCCAGTGTCCATGAAGGCA 1871
DB 361 CAGAAATTCAGATCCCAATTCACCTTGGAGCAGAGGCCAGTGTCCATGAAGGCA 420
QY 1872 AAAAAGAACCAATGCAAGTGTGTTTCTATCCAGAAAAATACAGGAACAGAGGAACAA 1931
DB 421 AAAAAGAACCAATGCAAGTGTGTTTCTATCCAGAAAAATACAGGAACAGAGGAACAA 480
QY 1932 AGCAGGATGACCTGAATCTGGATTATGGGTGAAGAGAGTACACACTAGTTCAG 1991
DB 481 AGCAGGATGACCTGAATCTGGATTATGGGTGAAGAGAGTACACACTAGTTCAG 540
QY 1992 TTTTCTCTTAACACGTGCCAAGCCAGGACGAGTCTTCCCTATGATACAGATTTCTT 2051
DB 541 TTTTCTCTTAACACGTGCCAAGCCAGGACGAGTCTTCCCTATGATACAGATTTCTT 600
QY 2052 TTGTCCTTGTCCATTAACCCCAAGACTTTTCTTAGATATATCTCTCACTATCCGTTATTC 2111
DB 601 TTGTCCTTGTCCATTAACCCCAAGACTTTTCTTAGATATATCTCTCACTATCCGTTATTC 660
QY 2112 AACCTTAGCTCTGCTTCTATTAATTTTAGGCTTTAGTATATATCTAATAGTTTGGCTT 2171
DB 661 AACCTTAGCTCTGCTTCTATTAATTTTAGGCTTTAGTATATATCTAATAGTTTGGCTT 719
QY 2172 TTGATGTGGATGATGTGAGCTTCATGTCTTAAATCTACTACAAGCATACCTTAACAT 2231
DB 720 TGAATGTGGATGATGTGAGCTTCATGTCTTAAATCTACTACAAGCATACCTTAACAT 779
QY 2232 GGTGATCTGCAA-GTAGTAGGCACCCCAATA 2260
DB 780 GGTGATCTGCAAGGTAGTAGGCACCCCAATA 809
```

RESULT 5
BI838266
LOCUS
DEFINITION
VERSION
ACCESSION
KEYWORDS
SOURCE
ORGANISM

BI838266
603083189Fl NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5222302 5',
mRNA sequence.
BI838266
BI838266.1 GI:15949816
human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL1559 row: a column: 23
High quality sequence stop: 770.

COMMENT

JOURNAL

AUTHORS

REFERENCE

FEATURES

source

Location/Qualifiers

1..777

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5222302"

/lab_host="DH10B"

/note="Organ: pooled pancreas and spleen; Vector:
pcMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."

BASE COUNT 216 a 169 c 177 g 215 t

ORIGIN

Query Match 30.0%; Score 732.4; DB 13; Length 777;

Best Local Similarity 97.2%; Pred. No. 1.3e-173;

Matches 756; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

QY 636 AAAAAGAGTCAAAAGAGAGGATTTTATGAAGATCTTGACAGATTTGAGAAAATGGTA 695

DB 1 AAAAAGAGTCAAAAGAGAGGATTTTATGAAGATC-TGACAGATTTGAGAAAATGGTA 59

QY 696 CCCAGGATCAGGATCAGCCCATATACATTCGCAAGCTTTCCCTTCATATAATAT 755

DB 60 CCCAGGATCAGGATCAGCCCATATACATTCGCAAGCTTTCCCTTCATATAATAT 119

QY 756 TTGACCGGACCTAGTGTCTACTGAGTGGCAAGTCTATACAGATTCCTCCCGCAG 815

DB 120 TTGACCGGACCTAGTGTCTACTGAGTGGCAAGTCTATACAGATTCCTCCCGCAG 179

QY 816 TCCAGCCTCGGAAATTCGACCCCTTCTGCTCTTCTCGCTGCTGCTCATATATGATA 875

DB 180 TCCAGCCTCGGAAATTCGACCCCTTCTGCTCTTCTCGCTGCTGCTCATATATGATA 239

QY 876 TTAGTTTCCATCGGATTCCTTCTCACAATCAATCTGTTTTTGTATGAGAACAGGAG 935

DB 240 TTAGTTTCCATCGGATTCCTTCTCACAATCAATCTGTTTTTGTATGAGAACAGGAG 299

QY 936 GATTTGATGTGGAGAAATAGATGTGAGGATGAAGTACTGCTGGACTGAGATCAGCT 995

```

Db      300 GATTGTGGATGGAGAAATAGAAATGTAGAGATGAAGTGAAGTGGAGTGAAGATCAGCT 359
QY      996 GCTTACGCTCAAGGGTCAAAATGATCTACTTACCTGAAGCAGATAGATCACTTTTCTAT 1055
Db      360 GCTTACGCTCAAGGGTCAAAATGATCTACTTACCTGAAGCAGATAGATCACTTTTCTAT 419
QY      1056 GTTACCAAGTGTATGAACCTGGAGCATTTGACAAGGAGAGGGCTGTATCTAAGTGACA 1115
Db      420 GTTACCAAGTGTATGAACCTGGAGCATTTGACAAGGAGAGGGCTGTATCTAAGTGACA 479
QY      1116 TCCCTCTGATGATGCCCGCGCATCTTGTCTTTTGGGAGACAATTTAGAGGAAT 1175
Db      480 TCCCTCTGATGATGCCCGCGCATCTTGTCTTTTGGGAGACAATTTAGAGGAAT 539
QY      1176 ACAAACTCAACCAAGAACTGGAAATCCTCAGTCAGAGGCTACAGCTCAGCTTAAAGAGCCC 1235
Db      540 ACAAACTCAACCAAGAACTGGAAATCCTCAGTCAGAGGCTACAGCTCAGCTTAAAGAGCCC 599
QY      1236 TGAAGATGAAAGAAAAGACAGACATGTGCTGATTTCTGTCCTTCCGTCGTTG 1295
Db      600 TGAAGATGACAAAGAAACAGACACATGTGCTGATTTCTGTCCTTCCGTCGTTG 659
QY      1296 CCAATGAGCTGGGCAACAGGCTCCAGTCCCTGCCAAAGATATGACAATGTGACCATCC 1355
Db      660 CCAATGAGCTGGGCAACAGGCTCCAGTCCCTGCCAAAGATATGACAATGTGACCATCC 719
QY      1356 TCTTTAGTGGCATTTGGGCTTCAATGCTTCTGTAGCAAGCATGCTGGAGAGG 1413
Db      720 TCTTTAGTGGCATTTGGGCTTCAATGCTTCTGTAGCAAGCATGCTGGAGCGG 777

```

RESULT 6

```

BG199621/c      823 bp      mRNA      linear      EST 21-APR-2001
LOCUS      RST18913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION      BG199621
ACCESSION      BG199621.1 GI:13721308
VERSION      EST.
KEYWORDS      human.
SOURCE      Homo sapiens

```

REFERENCE

```

AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 823)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.

```

Creation of genome-wide protein expression libraries using random

activation of gene expression

Nat. Biotechnol. 19 (5), 440-445 (2001)

2127151

Contact: Scott J. Cain

Athersys, Inc.

3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900

Fax: 216 361 9596

Email: scaine@atersys.com

High quality sequence stop: 348.

Location/Qualifiers

1..823

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Athersys RAGE Library"

/cell_line="Ht1080"

/note="See 'Creation of Genome-wide Protein Expression

Libraries using Random Activation of Gene Expression',

Nature Biotechnology, in press. Note that even though the

cell type indicated is Ht1080, since a random activation

method was used, these sequence tags are not necessarily

expressed in Ht1080 under normal circumstances."

249 a 165 c 146 g 263 t

BASE COUNT

ORIGIN

```

Query Match      29.8%; Score 728; DB 12; Length 823;
Best Local Similarity 96.4%; Pred. No. 1.6e-172;
Matches 787; Conservative 0; Mismatches 25; Indels 4; Gaps 4;
QY      1586 GCCTTGGACATGATGGAATCTGTCGCCAGGTTCAAGTAGATGGTGAATCTGTTCAGATA 1645
Db      821 GCCTTGGACATGATGGAATCTGTCGCCAGGTTCAAGTAGATGGTGAATCTGTTCAGATA 763
QY      1646 ACAATAGGATACACACTGGAGAGGTAGTACAGTGTCTATAGGACAGGGATGCTCGA 1705
Db      762 ACTATAGGATACACACTGGAGAGGTAGTACAGTGTCTATAGGACAGGGATGCTCGA 703
QY      1706 TACTGTCTTTTGGGAATCTGTCAACCTCACAAAGCCGACAGAAACACAGAGAGAAAG 1765
Db      702 TAATGTCTTTGGGAATCTGTCAACCTCACAAAGCCGATCAGAAACACAGAGAGAAAG 643
QY      1766 GGAATAAATGTGTCTGAATATACATACAGATGCTTATGTCTCCAGAAAATTCAGAT 1825
Db      642 GGAATAAATCAATGTCTCTGAATATACATACAGATGCTTATGTCTCCAGAAAATTCAGAT 583
QY      1826 CCACAAATCCACTTGGAGCACAGAGGCCAGTCCATGAAGGCAAAAAGAACCAATG 1885
Db      582 CCACAAATACACTTGGAGCACAGAGGCCAGTCCATGAAGGCAAAAAGAACCAATG 523
QY      1886 CAAGTTTGGTTTCTATCCAGAAAAATACAGGAACAGAGAAACAAAGCAGGATGATGAC 1945
Db      522 CAAGTTTGGTTTCTATCCAGAAAAATACAGGAACAGAGAAACAAAGCAGGATGATGAC 463
QY      1946 TGAATCTTGGATATGGGTGAAGAGGATACAGACTAGTTCACAGTTCCTTCCTAACAC 2005
Db      462 TGAATCTTGGATATGGGTGAAGAGGATACAGACTAGTTCCTTCCTAACAC 403
QY      2006 GTGCCAAGCCAGGAGCAGTCTTCCCTATGATACAGATTTCTTTCTCTGTCCTAT 2065
Db      402 GTGCCAAGCCAGGAGCAGTCTTCCCTATGATACAGATTTCTTTCTCTGTCCTAT 343
QY      2066 TACCCCAAGACTTTCTCTAGATATATCTCTCACTATCCCTATTCCTTATCAACCTTAGCTCTGC 2125
Db      342 TACCCCAAGACTTTCTCTAGATATATCTCTCACTATCCCTATTCCTTATCAACCTTAGCTCTGC 283
QY      2126 TTCTATATATCTTTAGGCTTTAGTATATATCTTAAAGTTTGGCTTTTGGTGGATGAT 2185
Db      282 TATCTATATCTTTTGGCTTTAGTATATATCTTAAAGTTTGGCTTTTGGTGGATAAT 223
QY      2186 GTGAGCTTCATGCTCTTAAATCTACTACAGCATTACCTAACATGGTGTGATCTGCACT 2245
Db      222 GTGAGCTTCATGCTCTTAAATCTACTACAGCATTACCTAACATGGTGTGATCTGCACT 163
QY      2246 AGTAGGACCCCAATAAATATTTTGAATTTAGTTAAATGAACTGAACAGTGTGTCG- 2304
Db      162 AGTAGGACCCCAATAAATATATATGTTGAATTTAGTTAAATGAACTGAACAGTGTGTCG 103
QY      2305 CATGTGTATATTTATATCATCTTTTACCAG-ATCTGTTTGTGTTTCCA-CATATATGATA 2362
Db      102 CATGTGTATATTTATATCATCTTTTACCAGATCTGTTTGTGTTTCCAACCATATATGATA 43
QY      2363 TGTATATTTTAAATGACTATAATGTTAAATGAAAGTTTAT 2398
Db      42 TGTATATTTTAAATGACTATAATGTTAAATGAAAGTTTAT 7

```

RESULT 7

```

BI666927      810 bp      mRNA      linear      EST 12-SEP-2001
LOCUS      603291676f1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:531146 5',
DEFINITION      mRNA sequence.
ACCESSION      BI666927
VERSION      BI666927.1 GI:15581160
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11788 row: g column: 19
 High quality sequence stop: 770.
 Location/Qualifiers
 1. 810
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5311146"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIMH/NHGRI, National
 Institutes of Health). Note: this is a NIH-MGC Library."
 244 a 168 c 175 g 223 t

BASE COUNT
 ORIGIN

Query Match 27.9%; Score 680.8; DB 13; Length 810;
 Best Local Similarity 97.4%; Pred. No. 1.3e-160;
 Matches 756; Conservative 0; Mismatches 12; Indels 8; Gaps 6;

QY 166 AAAAGAGCACAGTTAGATGAAGAGGACAGTTCTTCTGCAGATAATATATGATGACTC 225
 DB 39 AAAAGAGCACAGTTAGATGAAGAGGACAGTTCTTCTGCAGATAATATATGATGACTC 98
 QY 226 CAAACTATGATTTGGTGTGCTGCAAGCAAGTCTCAATCTCAATCTGCGAGAAAT 285
 DB 99 CAAACTATGATTTGGTGTGCTGCAAGCAAGTCTCAATCTCAATCTGCGAGAAAT 158
 QY 286 CCTCCAAATGTTGGGAGAGTGTCTTCTGCTTTTGGCAAGAACTGCTTATGATACAA 345
 DB 159 CCTCCAAATG-TTGGGAAGATGTTTCTGCTC-TTGGCAAGAACTGCTTATGATACAA 216
 QY 346 CTGCTGCTCTGGGCTCTAATGTGAGAGATTTCTACAGAACCTTGATGCTCTCAGCA 405
 DB 217 CTGCTGCTCTGGGCTCTAATGTGAGAGATTTCTACAGAACCTTGATGCTCTCAGCA 276
 QY 406 CCACCTTGCTACCACTACCCAGGAATCGGTCACCTCTCTTTAGGTCACCTGATGCGA 465
 DB 277 CCACCTTGCTACCACTACCCAGGAATCGGTCACCTCTCTTTAGGTCACCTGATGCGA 336
 QY 466 AAAGGGCAAGGACTCATTTTGCCTACTACTACAGAGAGAGGACCTTCAGGATATGCT 525
 DB 337 AAAGGGCAAGGACTCATTTTGCCTACTACTACTACAGAGAGAGGACCTTCAGGATATGCT 396
 QY 526 CATTTGGAATCATCAAAACAGTGGCCACAAATCCATCGCACTGGAATAGACATCAAGT 585
 DB 397 CATTTGGAATCATCAAAACAGTGGCCACAAATCCATCGCACTGGAATAGACATCAAGT 456
 QY 586 TATTCAGCAAGAAATGAAGATGTGATCATACTCAATTTTAAATGAAGAAAAAGATC 645
 DB 457 TATTCAGCAAGAAATGAAGATGTGATCATACTCAATTTTAAATGAAGAAAAAGATC 516

QY 646 AAAAGAAGAGGATTTTATGAGATCTTGACAGATTTGAAGAAATGTATCCAGGAATC 705
 DB 517 AAAAGAAGAGGATTTTATGAGATCTTGACAGATTTGAAGAAATGTATCCAGGAATC 576
 QY 706 ACGATACGCCCATATACATATCTGCAAGCTTTTCTTTCATATAATATTTGACCGGGA 765
 DB 577 ACGATACGCCCATATATATCTGCAAGCTTTTCTTTCATATAATATTTGACCGGGA 636
 QY 766 CCTAGTGTCTACTCAGTGTGGCAATCTATATACAGAGTTCTCCGCCAGCTCCAGCCCTGG 825
 DB 637 CCTAGTGTCTACTCAGTGTGGCAATCTATATACAGAGTTCTCCGCCAGCTCCAGCCCTGG 696
 QY 826 GAATTCGAGCC-TTCTGTCTGTCTCTCG--CTGTTTCTGCTCTCATATATGA-TATAGTT 881
 DB 697 GAATTCGAGCCCTTTCTGTCTGTCTCTCTGCTGCTGCTGCTCTCATATATGATTAAGATT 756
 QY 882 TCCATGGGATCCTTCTCACAATCAATCTGTTTGTATTGAGAGCAAGGAAGA 937
 DB 757 TCCATGGGATCCTTATATACATCTACTG-GTTGGATTGAGAGCAAGGAAGA 810

RESULT 8
 BI553448
 LOCUS 603193201r1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5264160 5',
 DEFINITION mRNA sequence.
 ACCESSION BI553448
 VERSION BI553448.1 GI:15440760
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 717)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11666 row: b column: 01
 High quality sequence stop: 712.
 Location/Qualifiers
 1. 717
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5264160"
 /clone_lib="NIH_MGC_95"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.5 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIMH/NHGRI, National
 Institutes of Health). Note: this is a NIH-MGC Library."
 207 a 157 c 172 g 181 t

BASE COUNT
 ORIGIN

Query Match 27.9%; Score 680.4; DB 13; Length 717;
 Best Local Similarity 98.3%; Pred. No. 1.6e-160;
 Matches 698; Conservative 0; Mismatches 11; Indels 1; Gaps 1;


```

Db      720 AGGCTCCA-TTTTCTCCTACACGTTCCAGCCGAGCACCA 759

RESULT 10
BO948975
LOCUS   939 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT 8879414 NCI_CGAP_Co24 Mus musculus cDNA clone
IMAGE:6475913 5', mRNA sequence.
ACCESSION BO948975
KEYWORDS EST.
SOURCE    house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 939)
AUTHORS   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL   NIH-MGC http://mgc.nci.nih.gov/.
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: The Cepko Laboratory
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: L14M14015 row: g column: 18
          High quality sequence stop: 658.
FEATURES
source
1..939
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6475913"
/lab_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 271 a 222 c 232 g 210 t 4 others
ORIGIN

Query Match 26.9%; Score 657.6; DB 14; Length 939;
Best Local Similarity 87.7%; Pred. NO. 9.1e-155;
Matches 739; Conservative 0; Mismatches 102; Indels 2; Gaps 2;

QY 1017 TGATCTACTTACCTGAAGCAGATAGCATACATCTTTTCTATGTTCACCAAGTGCATGAACC 1076
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1077 TGGAGGATTTGACAGGAGAGGGCTGTATCTAAGTGACATCCCTCTGCATGATGCCACGC 1136
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1137 GCGATCTTGTCTTTTGGGAGAACAAATTTAGAGAGGAATACAAATCCACCAAGAACATGG 1196
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 132 GAGACCTGGTCTTTTGGGAGAACAGTTCCGGGAGGAGTACAAATCTGACACAGAGCTGG 191
QY 1197 AATCTCTACCTGACAGGCTACAGCTCAGCTTAAGAGCCCTTGGAGATGAAGAAAGAAAAGA 1256
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 192 AATCTCTACCCGACAGGCTGACGCTCACACTGAGAGCCTTGGAGGATGAGAAAGAAAAGA 251
QY 1257 CAGACACATGCTGTATCTCTCCCTTCCCTGCTGTTGCCAATGAGCTGGCGCACAGC 1316
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 252 CAGACACATGCTGTATCTCTCCCTTCCCTGCTGTTGCCAATGAGCTGGAGACACAAGC 311
QY 1317 GTCCAGTGCCTGCCAAAAGATGACAAATGTGACCATCTCTTTTGTAGTGGCATTTGGGCT 1376
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 312 GCCCAGTGCCTGCCAAAAGATGACAAATGTGACCATCTCTTTCAGCGGCATTGGGCT 371

```

```

QY 1377 TCAATGCTTTCTGTAGCAAGCATGCAATCTGGAGAGGAGCCATGAAGATGCTCAACCTCC 1436
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 372 TCAATGCTTTCTGTAGCAAGCATGCAATCTGGAGAGGAGCCATGAAGATGCTCAATCTCC 431
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1437 TCAAGGACTCTACACACAGATTTGACACACATGACTGATTCGCGGAAAAACCATTTGTTT 1496
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 432 TCAAGGATCTCTACACCCGATTTGACACACATGACTGATTCAGGAAAAAACCATTTGTTT 491
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1497 ATAAGGTGGAGACTGTTGGTGACAAGTATATGACAGTGAAGTGGTTTACAGAGCCATGCA 1556
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 492 ACAAGGTGGAACAGTTGGTGACAAGTATATGACAGTGAAGTGGTTTACAGAGCCATGCA 551
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1557 TTCACCATGACAGTCCATCTGCCACCTGGCCCTTGGACATGATGAAGATGCTGCCACAG 1616
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 552 TCCACCATGACAGTCCATCTGCCACCTGGCCCTTGGACATGATGAAGATGCTGCCACAG 611
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1617 TTCAAGTAGATGGTGAATCTGTTCAAGTAAATAAGGATACACACTGGAGAGGTAGTTA 1676
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 612 TTCAAGTAGATGGTGAATCTGTTCAAGTAAATAAGGATACACACTGGAGAGGTAGTTA 671
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1677 CAGGTGTCATAGGACAGCGGATGCTGATCTGTTTGGGAAATACTGTCAACCTCA 1736
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 672 CAGGTGTCATAGGACAGCGGATGCTGATCTGTTTGGGAAATACTGTCAACCTCA 730
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1737 CAAGCCGACAGAAACACACAGGAGAAAGGAAAAATAAATGTGTCTGAATATACATAC- 1795
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 731 CAAGCAGGACAGAAACACACAGGAGAAAGGAAAAATAAATGTGTCTGAATATACATACA 790
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1796 AGATGTCCTATGTCTCCAGAAAAATTCAGATCCACAATTCCTGAGCAGACAGAGGCCCA 1855
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 791 AGGTGTCCTATGTCTCCAGAAAAATTCAGATCCACAATTCCTGAGCAGACAGAGGCCCA 1855
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 1856 GTG 1858
Db      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
QY 851 ANG 853

RESULT 11
BF036708 749 bp mRNA linear EST 20-OCT-2000
LOCUS 60145903F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3863401 5',
DEFINITION mRNA sequence.
ACCESSION BF036708
VERSION BF036708.1 GI:10744768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 749)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: DCTD/DP
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: L14M14015 row: e column: 02
          High quality sequence stop: 684.
FEATURES
source
1..749
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3863401"
/lab_lib="NIH_MGC_66"
/lab_host="adenocarcinoma"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;

```


QY 706 AGCATGAGCCCATATACATTCTGCAAGACGTTTCCCTTTTCATATATATTTGACCGGGA 765
 |||||
 Db 631 AGCATGAGCCCATATACATTCTGCAAGACGTTTCCCTTTTCATATATATTTGACCGGGA 690
 |||||
 QY 766 CCTAGTGGTCACTCAGTGTGCAATGCTATA 796
 |||||
 Db 691 CCTAGTGGTCACTCAGTGTGCAATGCTATA 721
 |||||

RESULT 13
 BG434435 808 bp mRNA linear EST 14-MAR-2001
 LOCUS 60250645F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4603908 5',
 mRNA sequence.
 ACCESSION BG434435
 VERSION BG434435.1 GI:13340941
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 808)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1CM1338 row: k column: 13
 High quality sequence stop: 698.

FEATURES
 source
 1. 808
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4603908"
 /clone_lib="NIH_MGC_79"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
 Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcgattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGGCGCGACATG-dn(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 Library."

BASE COUNT 242 a 179 c 181 g 206 t
 ORIGIN

Query Match 25.4%; Score 621; DB 12; Length 808;
 Best Local Similarity 94.5%; Pred. No. 1.5e-145;
 Matches 763; Conservative 0; Mismatches 30; Indels 14; Gaps 11;

QY 1370 GTGGGTTCAATGCTTTCTGTAGCAAGCATGCATCTGGGAAGGAGCGCATGAAGATCGTC 1429
 |||||
 Db 1 GTGGGTTCAATGCTTTCTGTAGCAAGCATGCATCTGGGAAGGAGCGCATGAAGATCGTC 60
 |||||

QY 1430 AACCTCTCAGGACCTTACACAGATTTGACACACTGACTGATTCGCCGGAACCA 1489
 |||||
 Db 61 AACCTCTCAGGACCTTACACAGATTTGACACACTGACTGATTCGCCGGAACCA 120
 |||||

QY 1490 TTTGTTTAAAGTGGAGACTGTGGTGACAAAGTATATGACAGTGAAGTGTATACACAG 1549
 |||||
 Db 121 TTTGTTTAAAGTGGAGACTGTGGTGACAAAGTATATGACAGTGAAGTGTATACACAG 180
 |||||

QY 1550 CCATGCATTCACCATGCACGATCCATCTGCCACCTGGCCTTGGACATGATGAATTTGCT 1609
 |||||
 Db 181 CCATGCATTCACCATGCACGATCCATCTGCCACCTGGCCTTGGACATGATGAATTTGCT 240
 |||||

QY 1610 GGCAGGTTCAAGTAGATGGTGAATCTCTTCAGATAACAATAGGATACACACTGGAG 1669
 |||||
 Db 241 GGCAGGTTCAAGTAGATGGTGAATCTCTTCAGATAACAATAGGATACACACTGGAG 300
 |||||

QY 1670 GTAGTTACAGGTGCATAGGACAGGGGATGCTCCATCTGTCTTTTGGGAATACTGTC 1729
 |||||
 Db 301 GTAGTTACAGGTGCATAGGACAGGGGATGCTCCATCTGTCTTTTGGGAATACTGTC 359
 |||||

QY 1730 AACCTCACAAGCCGAAACAGAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1788
 |||||
 Db 360 AACCTCACAAGCCGAAACAGAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
 |||||

QY 1789 TACATACAGATGTTCTTATGTCTCCAGAAAATTCAGATCCACAAATTCACACTGGAGCAG 1848
 |||||
 Db 420 TACATACAGATGTTCTTATGTCTCCAGAAAATTCAGATCCACAAATTCACACTGGAGCAG 479
 |||||

QY 1849 AGGCCAGTGTCCATGAAGGCG-AAAAAAGAACCAATGCAAGTTTGGTTCTTATCCAGAA 1907
 |||||
 Db 480 AGGCCAGTGTCCATGAAGGCGAAAAAAGAACCAATGCAAGTTTGGTTCTTATCCAGAA 539
 |||||

QY 1908 AAAATACAG 1967
 |||||
 Db 540 AAAATACAG 599
 |||||

QY 1968 AGAGGAG-TACAGACTAGTTCCAGTTTTCCTT-AAACAGTGCCCAAGCCAGGAGCAGT 2025
 |||||
 Db 600 AGAGGAGTTACAGACTAGTTCCAGTTTTCCTT-AAACAGTGCCCAAGCCAGGAGCAGT 659
 |||||

QY 2026 TCTTCCCTATGATGATACAGATTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2085
 |||||
 Db 660 TCTT-CTTATGATGATACAGATTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 715
 |||||

QY 2086 GATATATCTCTACATCCGTTATTCACCTTAGCTCTGCTTCTTCTTCTTCTTCTTCTTCTT 2145
 |||||
 Db 716 GATTTTCTCTCA-TATTCTGCTATGCAACCTTA-ATCTGCTCTCTCTCTCTCTCTCTCT 770
 |||||

QY 2146 TTAGTATATATCTAAAGTTGGGCTTT 2172
 |||||
 Db 771 TTAGTATATGATCTACAAAGTTGGGCTCT 797
 |||||

RESULT 14
 BM947644 751 bp mRNA linear EST 14-MAR-2002
 LOCUS UI-M-EG0p-bvb-n-08-0-UI.r1 NIH_BMAP_EG0p Mus musculus cDNA clone
 DEFINITION IMAGE:5689927 5', mRNA sequence.
 ACCESSION BM947644
 VERSION BM947644.1 GI:19431231
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 751)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)


```

FEATURES
  Seq primer: pyx-5.
  Location/Qualifiers
    1..751
      /organism="Mus musculus"
      /strain="C57BL/6"
      /db_xref="taxon:10090"
      /clone="IMAGE:568927"
      /clone_lib="NIH_BMAP_EG0p"
      /tissue_type="whole brain"
      /dev_stage="embryo 18.5 dpc"
      /lab_host="DH10B (T1 phage resistant)"
      /note="Organ: brain; Vector: pyx-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCAGCAGC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."
  BASE COUNT      210 a 176 c 182 g 182 t
  ORIGIN
    Query Match      25.3%; Score 618.8; DB 14; Length 751;
    Best Local Similarity 88.9%; Pred. No. 5.4e-145;
    Matches 668; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

  QY 879 GTTCCATGGATGCTTTCTCATCAATPACTGTTTGTATGAGAACAGGAGGAT 938
  Db 1 GTTCCATGGATGCTTTCTCATCAATPACTGTTTGTATGAGAACAGGAGGAT 60
  QY 939 TGTGGATGTGAGAAATAGAAATGTGAGGATGAATGACTGGGACTGAGATCAGTGT 998
  Db 61 TGTGGATGTGAGAAATAGAAATGTGAGGATGAATGACTGGGACTGAGATCAGTGT 120
  QY 999 TAGCTCTCAAGGTCAATGATCTACTTACCTGAACGATAGCATATCTTTCTATGTT 1058
  Db 121 TAGCTCTCAAGGCCAATGATCTACTTACGAGACGATAGCATCTCTCTCTGTT 180
  QY 1059 CACCAAGTCTATGAACCTGGACGATTTGACAAGGAGAGGGCTGTATCTTAAGTGACATCC 1118
  Db 181 CACCAAGTCTATGAACCTGGACGACCTAAACAAGAAGAGGCGCTGTATCTGAGTGACATCC 240
  QY 1119 CTCTGGATGATGCACGCCGCGATCTGTTCTTTTGGGAGAACAAATTTAGAGAGGAATACA 1178
  Db 241 CTCTCAGCATGCTACCCGAGACCTGGTCTTTTGGGAGAACAGATTCCGGGAGGAGTACA 300
  QY 1179 AACTCACCACAACTGGAAATCCCTACCTGACAGCTACAGCTCAGCTTAAGACCCCTGG 1238
  Db 301 AACTGACACAGAGCTGGAAATCCCTACCGCAGGCTGAGCTCAGCTGAGAGCCCTGG 360
  QY 1239 AAGATGAAAGAAAAGACAGACATGCTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1298
  Db 361 AGGATGAGAAAGAAAAGACAGACATGCTGTATCTCTCTCTCTCTCTCTCTCTCTCTCT 420
  QY 1299 ATGAGCTGGCGGACAAAGCCTCAGTGCCTGCCAAAAGATATGACAAATGACATCCTCT 1358
  Db 421 ATGAGCTGAGACAAAGCCTCAGTGCCTGCCAAAAGATATGACAAATGACATCCTCTCT 480
  QY 1359 TTATGGGCAATCTGGGCTTCAATGCTTCTGTAGCAAGCATGCATCTGGAGAGGAGCA 1418
  Db 481 TCAGCGGCAATCTGGGCTTCAATGCTTCTGTAGCAAGCATGCATCTGGAGAGGAGGCA 540
  QY 1419 TGAAGATCTCAACCTCTCAACGACCTCTACACCCAGATTTGACACACTGACTGATTC 1478
  Db 541 TGAAGATCTCAACCTCTCAACGATCTCTACACCCGATTTGACACACTGACTGATTCAC 600

```

```

1479 GGAAGAACCCATTGTTTATAAGGTGGAGACACTGTTGGTGACAAGTATATGACAGTGAGTG 1538
  Db 601 GGAAGAACCCATTGTTTATAAGGTGGAGAACAGTGTGTTGACAAGTATATGACAGTGAGTG 660
  QY 1539 GTTTACCCAGACCATTCATCCATCCAGCAGCATCCATCTGCCACCTGGGCTTGGACATGA 1598
  Db 661 GCTTGCCAGAACCTTGTATCCACCATGCACGGTCCCATTTGCCCACTGGCTTTAGACATGA 720
  QY 1599 TGGAAATTCCTGCGCCAGGTTCAAGTAGATGG 1629
  Db 721 TGGAAATTCCTGCGTCAAGTTCANGTAGATGG 751

  RESULT 15
  BI668844
  LOCUS      745 bp mRNA linear EST 12-SEP-2001
  DEFINITION 603294809F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5314058 5',
  mRNA sequence.
  ACCESSION BI668844
  VERSION   BI668844.1 GI:15583077
  KEYWORDS  EST.
  SOURCE    human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 745)
  AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11796 row: a column: 03
            High quality sequence stop: 745.

  FEATURES
    Location/Qualifiers
      1..745
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5314058"
        /clone_lib="NIH_MGC_96"
        /tissue_type="hypothalamus"
        /lab_host="DH10B"
        /note="Organ: brain; Vector: pBluescriptR (modified
        pBluescript KS+); Site_1: BamHI; Site_2: Sali-XhoI (gtcgag
        ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
        size-selected for average insert size 2.3 kb and
        normalized to 5x. This is a primary library enriched
        for full-length clones and constructed using the
        Cap-trapper method (Carninci, in preparation). Library
        constructed by M. Brownstein (NIH/NHGRI, National
        Institutes of Health). Note: this is a NIH_MGC Library."
    BASE COUNT      217 a 167 c 165 g 196 t
    ORIGIN
      Query Match      25.0%; Score 611.6; DB 13; Length 745;
      Best Local Similarity 98.3%; Pred. No. 3.5e-143;
      Matches 639; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

  QY 166 AAAAGAGGCACAGTTAGATGAAGAGGACAGTTCCTGTCAACAATAATATGATGACATC 225
  Db 91 AAAAGAGGCACAGTTAGATGAAGAGGACAGTTCCTGTCAACAATAATATGATGACATC 150
  QY 226 CAAAACCTTATGATTTGGTTGCTGCTGCAAGCAAGTCCCAATCTCAATGCTGGAGAAAT 285
  Db 151 CAAAACCTTATGATTTGGTTGCTGCTGCAAGCAAGTCCCAATCTCAATGCTGGAGAAAT 210

```

QY		286	CCCTCCAATGTTGGGAAGATGTTTTTCGTCTTTTGGCAAGAATCTGGTTATGATACAAAT	345
Db		211	CCGCCAATGTTGGGAAGATGTTTTTCGTCTTTTGGCAAGAATCTGGTTATGATACAAAT	270
QY		346	CTTGCGGTGCTCGGCCTTAATGTCAGAGAATTTCTCAGAAACCTTGATGCTCTGCACGA	405
Db		271	CTTGCGGTGCTCGGCCTTAATGTCAGAGAATTTCTCAGAAACCTTGATGCTCTGCACGA	330
QY		406	CCACCTTGCTACCATCTACCCAGAAATGCGTGCACCTTCTTTTAGTGCACTGATGCAGA	465
Db		331	CCACCTTGCTACCATCTACCCAGAAATGCGTGCACCTTCTTTTAGTGCACTGATGCAGA	390
QY		466	AAAGGGCAAAGGACTCATTTTGGCACTACTACTCAGAGAGAGAAAGGACTTTCAGGATATTGTT	525
Db		391	AAAGGGCAAAGGACTCATTTTGGCACTACTACTCAGAGAGAGAAAGGACTTTCAGGATATTGTT	450
QY		526	CATTGGNAATCATCAAACAGTGGGCACAACAATAATCCATGGCACTGAAATAGACATGAAGGT	585
Db		451	CATTGGNAATCATCAAACAGTGGGCACAACAATAATCCATGGCACTGAAATAGACATGAAGGT	510
QY		586	TATTCACCAAGAANAATGAAGAATCTGATCATACTACTACTCAATTTTAAATTGAAGAAAAAGATC	645
Db		511	TATTCACCAAGAANAATGAAGAATCTGATCATACTACTCAATTTTAAATTGAAGAAAAAGATC	570
QY		646	AAAAGAAGAGGATTTTATGAAGATCTTGACAGATTTTGAAGAAA-ATGGTACCAGGAAT	704
Db		571	AAAAGAAGAGGATTTTATGAAGATCTTGACAGATTTTGAAGAAAACATGGTACCAGGAAT	630
QY		705	CAGGCATCAGCCCCATATACATCTCGAAAAGCTTTTCCCTTTTCATATATATTGTGACCCGG	764
Db		631	CAGGCATCAGCCCCATATACATCTCGAAAAGCTTTTCCCTTTTCATATATATTGTGACCCGG	690
QY		765	ACCTAGTGGTCACTCACTGTGGCAATGCTATPATACAGAGTTCTCCCCCAG	814
Db		691	A-CTAGTGGTCACTCACTGTGGCACTGTATATATACAGAGTTCTCCCCCAG	739

Search completed: July 1, 2003, 17:26:06
Job time : 2242.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 13:06:07 ; Search time 81.9108 Seconds
(without alignments)
9146.676 Million cell updates/sec

Title: US-09-762-767A-3
Perfect score: 2443
Sequence: 1 cccccccgcgcgtgcgc.....aaatcattttctaaaggagt 2443

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81.4	3.3	3603	3 US-08-908-643C-85	Sequence 85, Appl
2	81.4	3.3	3645	3 US-08-908-643C-83	Sequence 83, Appl
3	81.4	3.3	3745	3 US-08-908-643C-84	Sequence 84, Appl
4	81.4	3.3	3787	3 US-08-908-643C-82	Sequence 82, Appl
5	60.2	2.5	7218	1 US-08-232-463-14	Sequence 14, Appl
6	56.4	2.3	4533	3 US-08-726-214-5	Sequence 5, Appl1
7	54.8	2.2	3784	1 US-07-623-033-1	Sequence 1, Appl1
8	53.6	2.2	3357	3 US-08-726-214-7	Sequence 7, Appl1
9	52.4	2.1	5199	3 US-08-726-214-13	Sequence 13, Appl
10	52.2	2.1	3518	4 US-09-412-210-2	Sequence 2, Appl1
11	50.8	2.1	4131	3 US-08-726-214-11	Sequence 11, Appl
12	50.6	2.1	1652	3 US-08-726-214-17	Sequence 17, Appl
13	50.6	2.1	4008	3 US-08-307-896-5	Sequence 5, Appl1
14	50.6	2.1	4008	3 US-08-726-214-3	Sequence 3, Appl1
15	50.6	2.1	4008	5 PCT-US95-11808-5	Sequence 5, Appl1
16	49.2	2.0	3924	3 US-08-726-214-9	Sequence 9, Appl1
17	48.6	2.0	2092	3 US-08-307-896-6	Sequence 6, Appl1
18	48.6	2.0	2092	5 PCT-US95-11808-6	Sequence 6, Appl1
19	48.6	2.0	4601	3 US-08-726-214-15	Sequence 15, Appl
20	47.4	1.9	7218	1 US-08-232-463-14	Sequence 14, Appl
21	46.2	1.9	3549	4 US-09-008-097-5	Sequence 5, Appl1
22	46.2	1.9	4342	4 US-09-474-076-1	Sequence 1, Appl1
23	46	1.9	4523	4 US-09-473-716-1	Sequence 1, Appl1
24	45.8	1.9	4046	1 US-07-793-961A-1	Sequence 1, Appl1
25	45.8	1.9	4046	1 US-08-240-357-1	Sequence 1, Appl1
26	42.6	1.7	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
27	42.6	1.7	4411529	4 US-09-103-840A-1	Sequence 1, Appl1

C	28	42.4	1.7	19124	2 US-08-487-826B-13	Sequence 13, Appl
	29	42.2	1.7	658	4 US-08-998-416-595	Sequence 595, App
	30	40.4	1.7	1440	4 US-09-134-001C-1922	Sequence 1922, Ap
	31	40	1.6	51952	3 US-08-947-823-1	Sequence 1, Appl1
	32	39.6	1.6	747	4 US-09-134-001C-2248	Sequence 2248, Ap
	33	39.4	1.6	872	4 US-08-998-416-487	Sequence 487, App
	34	39.4	1.6	3701	4 US-08-845-258-10	Sequence 10, Appl
	35	39.4	1.6	3701	4 US-08-990-571-10	Sequence 10, Appl
	36	39.4	1.6	3701	4 US-08-723-142A-10	Sequence 10, Appl
	37	39.4	1.6	3701	4 US-08-528-784A-10	Sequence 10, Appl
	38	39.4	1.6	72604	4 US-09-268-992-7	Sequence 7, Appl1
	39	39.4	1.6	72604	4 US-09-657-474-7	Sequence 7, Appl1
	40	38.8	1.6	1422	1 US-08-319-704-5	Sequence 5, Appl1
	41	38.6	1.6	4473	3 US-08-894-173-1	Sequence 1, Appl1
	42	38.6	1.6	4473	4 US-09-398-193-1	Sequence 1, Appl1
	43	38.6	1.6	5361	4 US-08-973-462-2	Sequence 2, Appl1
	44	38.6	1.6	6152	4 US-08-973-462-1	Sequence 1, Appl1
	45	38.4	1.6	4946	3 US-08-817-188-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-908-643C-85
; Sequence 85, Application US/08908643C
; Patent No. 6120995
; GENERAL INFORMATION:
; APPLICANT: Waldman, Scott A.
; Pearlman, Joshua M.
; Barber, Michael T.
; Schultz, Stephanie
; Parkinson, Scott J.
; TITLE OF INVENTION: COMPOSITIONS THAT SPECIFICALLY BIND TO
; COLORECTAL CANCER CELLS AND METHODS OF
; USING THE SAME
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6120995ris LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; FILING DATE: 07-Aug-1997
; CLASSIFICATION: N/A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark Deluca
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-2209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3603
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-08-908-643C-85

Query Match 3.3% ; Score 81.4 ; DB 3 ; Length 3603;

	Best Local Similarity	52.4%;	Pred. No. 3.2e-13;		Mismatches	236;	Gaps	4;
QY	1239	AAGATGAAAGAAAAAGACAGACACATTCGTCTATTCTCTCCTCCGCTCGTTGCCA	1298					
Db	2266	AGGCAGAGAGGGACAGGGCTGACAGACTTAACTTTATGTGCTTCCAAGGCTAGTGATA	2325					
QY	1299	ATGAGCTGCGGCACACAGCGTCAGTGCCTGCCAAAAGATATGCAATGTGACCATCCTCT	1358					
Db	2326	AGTCTCTGAAGGAGAAGGCTTTGTGGAGCCGGAACATATAGGAAGTTACAATCTACT	2385					
QY	1359	TTAGTGGCATTTGGGCTTCAAATGCTTCTGTAGCAAGCATGCATCTGGAGAAGGAGCCA	1418					
Db	2386	TCAGTGACATTTAGGTTTTCACTACTATCTGCAAAATACAGCACCC-----CCA	2433					
QY	1419	TGAAGATCGTCAACCTCCCTCAGGACCTCTACACAGATTTTGACACACTGACTGATCCC	1478					
Db	2434	TGGAAGTGTGGACATGCTTTAATGACATCTAAGAGTTTGTGACCACATTTGTGATCATC	2493					
QY	1479	GGAAAAACCCATTTGTTTATAAGGTGGAGACTGTTGGTGACAAGTATATGACAGTCAGTG	1538					
Db	2494	ATGA-----TGCTACAAGGTGAAACCCATCGTGATCGGTACATGGTGGCTAGTG	2544					
QY	1539	GTTTACCAGACCA----TGCATTCACATGACGAGATCCATCTGCCACCTGGCCTTGGAA--	1593					
Db	2545	GTTTGGCTTAAGAGAAATGGCAATCGGCATGCAATGACATTTGCAAGATGGCCTTGGAAA	2604					
QY	1594	-----CATGATGGAATTTGCTGGCCAGGTTCAACTAGATGGTGAATCTGTTTCAGATAA	1646					
Db	2605	TCCTCAGCTTCATGGGACCTTTGAGCTGGAGCATCTTCTTGGCCTCCCAATATGATTC	2664					
QY	1647	CRAATAGGGATACACACTGAGAGGTTAGTTACAGGTGTCATAGGACAGCGGATGCCCTCGAT	1706					
Db	2665	GCATTGGAGTTCACTTGGTCCCTGTGCTGCTGGAGTTTGTGGGAATCAAGATGCCCTCGTT	2724					
QY	1707	ACTGTCTTTTGGGAATCTGTCAACCTCCACAAGCCGAACAGAAACACACAGGAGAAAGS	1766					
Db	2725	ATTGCTATTTTGGAGATACGGTCAACACAGCCCTCTAGGATGGNAATCCATGGCCTCCCTT	2784					
QY	1767	GA AAAATAAATGTGCTGAATATAC	1791					
Db	2785	TGAGATTTCAACGTGAGTGGCTCCAC	2809					

RESULT 2

US-08-908-643C-83
; Sequence 83, Application US/08908643C
; Patent No. 6120995

APPLICANT: Waldman, Scott A.

FAIRLAWSON, SCOTT U.
TITLE OF INVENTION: COMPOSITIONS THAT SPECIFICALLY BIND TO
COLORECTAL CANCER CELLS AND METHODS OF
USING THE SAME

NUMBER OF SEQUENCES: 85

NUMBER OF SEQUENCES: 03
CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6120995ris LLP
STREET: One Liberty Place - 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch

COMPUTER: IBM PC compatible

RESULT 3

US-08-908-643C-84

03 08 208 043C 04
: Sequence 84, Application US/08908643C

; Patent No. 6120995

; RECORD NO. 012055
; GENERAL INFORMATION:

APPLICANT: Waldman, Scott A.


```
QY 1359 TTAGTGGCATTTGGGCTTCAATGCTTTCTGTAGCAAGCATGTCATCTGGAGAGGAGCCA 1418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2597 TCAGTGACATTTGAGTGTTCACCTACTATCTGCAATATACGACACCC-----CCA 2644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1419 TGAAGATCGTCACCTCCTCAGGACCTCTACACAGATTTGACACACTGATGATCC 1478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2645 TGAAGTGTGGACATGCTTAATGACATCTATAAGAGATTTTGACCAACATTTGTTGATCATC 2704
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1479 GGAAGAACCCATTTGTTTATAGGTGGAGACTGTTGTCACAGATATATGACAGTCAAGTG 1538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2705 ATGA-----TGCTACAGGTGGAACCAACATCGGTGATGCTGATGCTGGCTAGTG 2755
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1539 GTTACAGAGCCA---TGCAATTCACCATGCAAGATCCATCTGCCACCTGGCCTTGGA-- 1593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2756 GTTTCCTAAGAGAAATGCAATCGCATGCAATAGACATTCGCAAGATGGCCTTGGA 2815
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1594 -----CATGATGGAATGCTGGCCAGGTTCAGTAGATGTTGATCTGTTTCAGATAA 1646
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2816 TCTCAGCTTCATGGGGACCTTTGAGCTGGAGCATCTTCTGGCCCTCCCAATATGATGATTC 2875
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1647 CAATAGGATACACACTGGAGAGGTAGTTACAGGTGTATAGACAGCGGATGCCCTCGAT 1706
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2876 GCATTGGAGTTACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2935
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1707 ACTGCTTTTGGGAATCTGTCAACCTCACAAGCCGACGACGAAACCAACAGGAGAAAGG 1766
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2936 ATTGCTATTGGAGATACGGTCAACACAGCCCTTAGGATGGAATCCACTGGCCTCCCTT 2995
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1767 GAAATAATATGCTGCTGAATATAC 1791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2996 TGAGAAATTCAGTGTAGTGGCTCCAC 3020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 5

```
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
```

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7218 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: PTZgpt-F1s

; US-08-232-463-14

; Query Match

; Best Local Similarity 5.68; Pred. No. 6.9e-07;

; Matches 23; Conservative 224; Mismatches 162; Indels 0; Gaps 0;

; QY 1620 AAGTAGATGCTGAATCTGTTTCAGATAACAATAGGATACACACTGGAGAGGTAGTTACAG 1679

; Db 1431 RRR 1372

; QY 1680 GTGTATAGGACAGCGGATGCTCGATCTCTTTTGGGAATCTGTCACCTCACAA 1739

; Db 1371 RRR 1312

; QY 1740 GCGGACAGAACACAGAGAGAAAGGAGAAATAAATGTCTGTAATATACATACAGAT 1799

; Db 1311 RRR 1252

; QY 1800 GTCTTATGCTCCAGAAAATTCAGATCCACAATTCACCTGGAGCAGAGCCCGAGTGT 1859

; Db 1251 RRR 1192

; QY 1860 CCATGAAGGCGAAAGAACCAATGCAAGTTGGTTTCTATCCAGAAAAATACAGGAA 1919

; Db 1191 RRR 1132

; QY 1920 CAGAGGAACAAGCAGGATGATGACTGAATCTTGGATTCGGGTGAAGAGAGTACAG 1979

; Db 1131 RRR 1072

; QY 1980 ACTAGTTCACAGTTCTCTTAACACAGCTGCCAAGCCAGCAGCAGCTTCT 2028

; Db 1071 RRRRRATCGCAAGCTCCCTCGACCTGCAGCCAGCTCGGAATTAATCT 1023

RESULT 6

```
US-08-726-214-5
; Sequence 5, Application US/08726214
; Patent No. 6107076
; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,214
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,498
; FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
```

```

; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:450
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
;
; INFORMATION FOR SEQ ID NO: 5:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4533 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-726-214-5

```

Query Match 2.3%; Score 56.4; DB 3; Length 4533;
Best Local Similarity 61.6%; Pred. No. 6.5e-06;
Matches 90; Conservative 0; Mismatches 56; Indels 0

```

RESULT 7
US-07-623-033-1
; Sequence 1, Application US/07623033
; Patent No. 5237051
; GENERAL INFORMATION:
; APPLICANT: Garbers, David L.
; APPLICANT: Schulz, Stephanie
; TITLE OF INVENTION: CLONING THE ENTEROTOXIN RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TILTON, FALLON, LONGMUS & CHESTNUT
; STREET: 100 South Wacker Drive - Suite 960
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/623,033
; FILING DATE: 19901206
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: VD9018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 456-8000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3784 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; IMMEDIATE SOURCE:
; LIBRARY: gtl1 cDNA
; CLONE: #11 A
; FEATURE:

```

[illegible]


```

; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/726,214
; APPLICATION NUMBER: 435
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-726-214-7

Query Match 2.2%; Score 53.6; DB 3; Length 3357;
Best Local Similarity 60.1%; Pred. No. 3.5e-05;
Matches 89; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1650 TAGGGATACACACTGGAGAGGTAGTTACAGGTGTCTATAGGACGCGGATGCTCGATATG 1709
Db 3030 TGGGGTTAAACCATGGACAGCTGCTAGCAGGGGTGATTGGGCGACAGAGCCCAATATG 3089

QY 1710 GTCTTTTGGGAATCTGTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 1769
Db 3090 ACATCTGGGGGAATCTGTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 3149

QY 1770 AAATAAATGTCTGTAATATACATACAG 1797
Db 3150 AGATCCAACTGACCGAGGAGACAGCTAG 3177

RESULT 9
US-08-726-214-13
; Sequence 13, Application US/08726214
; Patent No. 6107076
; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/726,214
; APPLICATION NUMBER: 435
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-726-214-7

Query Match 2.1%; Score 52.4; DB 3; Length 5199;
Best Local Similarity 56.3%; Pred. No. 0.0001;
Matches 98; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1648 AATAGGGATACACACTGGAGAGGTAGTTACAGGTGTCTATAGGACGCGGATGCTCGATA 1707
Db 3702 AGTCGGCATAAACACCGGCGCTGTGTTGCTGGAGTGTGGAGCAGCGCAAGCCCTCAGTA 3761

QY 1708 CTGTCTTTTGGGAATCTGTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 1767
Db 3762 TGACATCTGGGGAAACACAGTCAATGTTGCCAGCGCATGAGAGCAGCGGAGAGCTTGG 3821

QY 1768 AAAATAAATGTCTGTAATATACATACAGATGCTTATGTCTCCAGAAATTC 1821
Db 3822 GAAATCCAGGTACCGAAGACATGCACTATCTCCAGGACTCGGATATTC 3875

RESULT 10
US-09-412-210-2
; Sequence 2, Application US/09412210
; Patent No. 6403358
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 21529, A NOVEL ADENYLYL CYCLASE
; FILE REFERENCE: 5800-47
; CURRENT APPLICATION NUMBER: US/09/412,210
; CURRENT FILING DATE: 1999-10-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3518
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: 21529 adenylyl cyclase
; NAME/KEY: CDS
; LOCATION: (247)...(3480)
; US-09-412-210-2

Query Match 2.1%; Score 52.2; DB 4; Length 3518;
Best Local Similarity 60.0%; Pred. No. 9.1e-05;
Matches 87; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1648 AATAGGGATACACACTGGAGAGGTAGTTACAGGTGTCTATAGGACGCGGATGCTCGATA 1707
Db 3198 AGTGGGTTGAACCATGGACCGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 3257

QY 1708 CTGTCTTTTGGGAATCTGTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 1767
Db 3258 TGACATTTGGGGCAACACAGTGAACGTGGCCAGCGCATGAGAGTACAGGAGTCTCTTGG 3317

QY 1768 AAAATAAATGTCTGTAATATACATA 1792
Db 3318 CAAATCCAAAGTACTGAGGAGACA 3342
```

```

; CURRENT FILING DATE: 1994-09-16
;
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 4008
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-08-307-896-5

Query Match 2.1%; Score 50.6; DB 3;
Best Local Similarity 59.3%; Pred. No. 0.00029;
Matches 86; Conservative 0; Mismatches 59;

```

1648	QY	ATAAGGGATACACACTGGAGAGGTACTTACAGTGTTCATAGGACACGGGTGCTCGATA	1707
3060	Db	AGTGGGTATCAACCATGGGCTGTAAATAGTGGCGTCATAGGGGCTCAAAGGCCACAGTA	3119
1708	QY	CTGTCTTTTGGGAATACATGTCTCAACTCACAAGCCGAACAGAAACCACAGGAGAAAGGG	1767
3120	Db	TGACATCTGGGGCCACACACTGTCAACGTGGCCAGCAATGGACAGCACCGGGTCTCTGGA	3179
1768	QY	AAAAATAAATGTGTCTGAATATACA	1792
3180	Db	CAAAATACAGGTCACTGAGGAGACA	3204

```

RESULT 14
US-08-726-214-3
; Sequence 3, Application US/08726214
; Patent No. 6107076
;
; GENERAL INFORMATION:
;
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

Query Match	2.1%;	Score 50.6;	DB 3;	Length 4008;
Best Local Similarity	59.3%;	Pred. No. 0.00029;		
Matches	86;	Conservative 0;	Mismatches 59;	Indels 0; Gaps 0;
QY	1648	AAATGGGATACACACTGGAGAGGTAGTTACAGGTGTCATAGACACCGGATGCCTCGATA	1707	
Db	3060	ACTGGGTATCAACCATGGGCCCTGTAAATAGCTGGCTCATAGGGGCTCAAAGGCCACAGTA	3119	
QY	1708	CTGCTCTTTTGGGAATACTGTCAACCTCTCAAGCCGAAACAGAAACCCACAGGAGAAAGGG	1767	
Db	3120	TGACATCTGGGGCAACACTGTCAACGTGGCCAGCAATGGACAGCACCGGGGTCTCTGGA	3179	
QY	1768	AAAAATAAATGTGCTGAATATACA	1792	
Db	3180	CAAAATACAGGTGACTGAGGAGACA	3204	

RESULT 15
PCT-US95-11808-5
; Sequence 5, Application PC/TUS9511808
; GENERAL INFORMATION:
; APPLICANT: Iyengar, Srinivas Ravi V.
; TITLE OF INVENTION: MUTANT ACTIVATED GSALPHA AND
; TITLE OF INVENTION: ADENYLK
; TITLE OF INVENTION: CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue and
; ADDRESSEE: Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11808
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/307,896
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Richard S.
; REGISTRATION NUMBER: 26,154
; REFERENCE/DOCKET NUMBER: 29970 165/28755
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 408-2500
; TELEFAX: (212) 765-2519
; TELEX: 650 6111063
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4008 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
PCT-US95-11808-5

```

Query Match      2.1%; Score 50.6; DB 5; Length 4008;
Best Local Similarity 59.3%; Pred. No. 0.00029;
Matches 86; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY      1648 AATAGGGATACACACTGGAGAGGTAGTTACAGTGTCTATAGACACGCCGATGCTCTCGATA 1707
Db      3060 AGTGGGTATCAACCATGGCCCTGTTAATACGTGGCTCATAGGGCTCAAAAGCCACAGTA 3119

QY      1708 CTGTCCTTTTTGGGAATACTGTCAACCTCTACAGCCGCACAGAAACACACAGGAGAAAGGG 1767
Db      3120 TGACATCTGGGGCACACTGTCAACGTGGCCACAGAAATGGACAGCACCGGGCTCTCGGA 3179

QY      1768 AAAAATAAATGTCTGTAATATACA 1792
Db      3180 CAANAATACAGGTACTGGAGGACA 3204

Search completed: July 1, 2003, 17:29:29
Job time : 88.9108 secs

```

Result No.	Score	Query		ID	Description
		Match	Length		
1	570	23.3	1302	9	US-10-106-698-961
2	521	21.3	177556	9	US-09-952-213D-6
3	197.4	8.1	15093	9	US-09-952-213D-5
4	192.6	7.9	2850	9	US-10-205-823-163
5	192.6	7.9	2871	9	US-10-205-823-164
6	192.6	7.9	3017	9	US-10-205-823-161
7	187.2	7.7	2430	9	US-09-952-213D-1
8	147.2	6.0	2335	10	US-09-917-800A-1569
9	131.8	5.4	7697	9	US-09-952-213D-4
10	130.6	5.3	177556	9	US-09-952-213D-6
11	111.8	4.6	15093	9	US-09-952-213D-5
12	86	3.5	955	9	US-09-952-213D-3
13	81.4	3.3	3745	9	US-10-157-031-17
14	81.4	3.3	3787	10	US-09-819-249-1
15	56.4	2.3	7697	9	US-09-952-213D-4
16	52.6	2.2	1037	9	US-10-198-846-13184
17	52.6	2.2	6196	9	US-10-282-942-1
18	52.6	2.2	7053	10	US-09-764-864-792
19	52.4	2.1	634	9	US-10-198-846-9332

Db	181	GTCCCTGTGCTCAATTACCCCAAGACTTTCTTCCTAGATATATCTCTCACTATCCGGTTATTCA	240
Qy	2114	CCTTAGCTCTGCTTTCTATTACTATTTTAGGCTTTAGTATATATCTAAAGTTTGGCTTTT	2173
Db	241	CCTTAGCTCTGCTTTCTATTACTATTTTAGGCTTTAGTATATATCTAAAGTTTGGCTTTT	300
Qy	2174	GATGTGGATGATGTGAGCTTCATGTGCTCTAAATCTACTACAAGCATTAACCTAACATGG	2233
Db	301	GATGTGGATGATGTGAGCTTCATGTGCTCTAAATCTACTACAAGCATTAACCTAACATGG	360
Qy	2234	TGATCTGCAAGTAGTAGGACCCCAATAAATATTTGTTGAATTTAGTTAAATGAACCTGAA	2293
Db	361	TGATCTGCAAGTAGTAGGACCCCAATAAATATTTGTTGAATTTAGTTAAATGAACCTGAA	420
Qy	2294	CAGTGTGTGGCCATGTGTATATTTATATCATGTTTACCAAACTGTTTAGTGTCCACAT	2353
Db	421	CAGTGTGTGGCCATGTGTATATTTATATCATGTTTACCAAACTGTTTAGTGTCCACAT	480
Qy	2354	ATATGTATATGTATATTTTAAATGACATATATGTGTAATAAAGTTTATATCATGTGTGTAT	2413
Db	481	ATATGTATATGTATATTTTAAATGACATATATGTGTAATAAAGTTTATATCATGTGTGTAT	540
Qy	2414	ATCATTTATAGAAATCATTTTCTAAAGAGT	2443
Db	541	ATCATTTATAGAAATCATTTTCTAAAGAGT	570

```

RESULT 2
US-09-952-213D-6
; Sequence 6, Application US/09952213D
; Publication No. US20030096240A1
; GENERAL INFORMATION:
; APPLICANT: MURAD, FERID
; APPLICANT: SHARINA, IRAIDA G.
; APPLICANT: KRUMENACKER, J. S.
; APPLICANT: MARTIN, E.
; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SGC
; FILE REFERENCE: UTSH:252US
; CURRENT APPLICATION NUMBER: US/09/952,213D
; CURRENT FILING DATE: 2002-08-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 177556
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2293..144567)
; OTHER INFORMATION: N = A, C, T/U OR G
US-09-952-213D-6

```

Query Match	21.3%	Score 521;	DB 9;	Length 177556;
Best Local Similarity	100.0%;	Pred. No. 1e-126;		
Matches 521;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1923	AGGAACAAGCAGGATGATGACTGAATCTTGATATATGGGTGAGAGGATACAGACT	1982	
DB	43237	AGGAACAAGCAGGATGATGACTGAATCTTGATATATGGGTGAGAGGATACAGACT	43296	
QY	1983	AGGTTCCAGTTTCTCTAACACGTGCCAAGCCAGGACAGTCTTCTCCTATGGATACA	2042	
DB	43297	AGGTTCCAGTTTCTCTAACACGTGCCAAGCCAGGACAGTCTTCTCCTATGGATACA	43356	
QY	2043	GATTTTCTTTTGTCCTTGTCCTATACCCCAAGACTTTCTCTAGATATATCTCTCACTAT	2102	
DB	43357	GATTTTCTTTTGTCCTTGTCCTATACCCCAAGACTTTCTCTAGATATATCTCTCACTAT	43416	
QY	2103	CGGTTATCAACCTTAGCTCTGCTTTCTATTAATCTTTTAGGCTTAGTATATATCTAAA	2162	
DB	43417	CGGTTATCAACCTTAGCTCTGCTTTCTATTAATCTTTTAGGCTTAGTATATATCTAAA	43476	
QY	2163	GTTTGGGCTTTTGATGTGSGATGATGTGAGCTTCATGTGCTTAAAAATCTACTACAAAGCAIT	2222	

43477	Db		GTTTGGCTTTGATGTGGATGAATGAGGCTTCATGTGCTTAAAAATCTACTACAAGCATT	43536
2223	Qy		ACCTAACATGGTGATCTGCAAGTAGTAGGCACCCCAATAAATATTGTTGAAATTTAGTTAA	2282
43537	Db		ACCTAACATGGTGATCTGCAAGTAGTAGGCACCCCAATAAATATTGTTGAAATTTAGTTAA	43596
2283	Qy		ATGAAACTCAACAGAGTTTGGCCATCTGTAATATTATATCATGTTTTACCAATCTGTTTA	2342
43597	Db		ATGAAACTCAACAGAGTTTGGCCATCTGTAATATTATATCATGTTTTACCAATCTGTTTA	43656
2343	Qy		GGTTCACATATATGATCA	2402
43657	Db		GGTTCACATATATGATCA	43716
2403	Qy		TGTTGGTGATATCATATTAGAAAATCATTTCTCAAAGGAGT	2443
43717	Db		TGTTGGTGATATCATATTAGAAAATCATTTCTCAAAGGAGT	43757

RESULT 3
US-09-952-213D-5
; Sequence 5, Application US/09952213D
; Publication No. US20030096240A1
; GENERAL INFORMATION:
; APPLICANT: MURAD, FERID
; APPLICANT: SHARINA, IRAIDA G.
; APPLICANT: KROMENACKER, J. S.
; APPLICANT: MARTIN, E.
; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SGC
; FILE REFERENCE: UTSH:752US
; CURRENT APPLICATION NUMBER: US/09/952,213D
; CURRENT FILING DATE: 2002-08-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 15093
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1213)..(8441)
; OTHER INFORMATION: N = A, C, G, or T/U
US-09-952-213D-5

	Query Match	8.1%;	Score 197.4;	DB 9;	Length 15093;		
	Best Local Similarity	83.4%;	Pred. No. 3.8e-41;				
Matches	211;	Conservative	19;	Mismatches	20; Indels	3; Gaps	1;
QY	1253	AAGACAGACACATTTGCTGTATTCTGCCTTCCCTTCCTCGTGTTGCCAATGAGCTGCGGCAC	1312				
Db	8855	AATTCCACAGATTGGTGATTCTGTCTCCTTCCCATCTGTTGCCAATGAGCTGAGACAC	8914				
QY	1313	AAGCGTCCAAGTGCCTGCCRAAGAATATGACAATGTGNACCATCTCTTTTAGTGGCAATTTGTG	1372				
Db	8915	AAGCGCCCCAGTGCCTGCCAAAAGATACGAANTGTGACCATCTCTTTCACGGGCATTTGTG	8974				
QY	1373	GGCTTCAAATGCTTTCTGTAGCAAGCATGCATCTGGAGAAGGACCATGAAGATCGTCTCAAC	1432		:	:	:
Db	8975	GGCTTCAATGCTTTCTGTAGCAAGCATGCATCTGGAGAAGGACCATGSWAGATTTGTCAWK	9034		:	:	:
QY	1433	CTCCTCAACGACTCTTACACAGATTTGACACACTGACTGATTC --- CGGAAAAAACCCA	1489		:	:	:
Db	9035	CTCSTCARCGRYCKVANSWSYC GATNTGYAYACTGACNGAA TTCAGCGCGAAA AACCCA	9094		:	:	:
QY	1490	TTTTGTTTTATAAGG	1502				
Db	9095	TTTTGTTTTATAAGG	9107				

RESULT 4
US-10-205-823-163
: Sequence 163, Application US/10205823

Publication No. US20030108963A1					
; GENERAL INFORMATION:					
; APPLICANT: Schlegel, Robert					
; APPLICANT: Monahan, John E.					
; APPLICANT: Endege, Wilson O.					
; APPLICANT: Gannavarapu, Manjula					
; APPLICANT: Gorbacheva, Bella					
; APPLICANT: Hoersch, Sebastian					
; APPLICANT: Kamatkar, Shubhangi					
; APPLICANT: Wonsey, Angela M.					
; APPLICANT: Glatt, Karen					
; APPLICANT: Zhao, Xumei					
; APPLICANT: Anderson, Dustin					
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND					
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND					
; FILE REFERENCE: MRI-044					
; CURRENT APPLICATION NUMBER: US/10/205,823					
; PRIOR FILING DATE: 2002-07-25					
; PRIOR APPLICATION NUMBER: 60/307,982					
; PRIOR FILING DATE: 2001-07-25					
; PRIOR APPLICATION NUMBER: 60/314,356					
; PRIOR FILING DATE: 2001-08-22					
; PRIOR APPLICATION NUMBER: 60/325,020					
; PRIOR FILING DATE: 2001-09-25					
; PRIOR APPLICATION NUMBER: 60/341,746					
; PRIOR FILING DATE: 2001-12-12					
; PRIOR APPLICATION NUMBER: 60/362,158					
; PRIOR FILING DATE: 2002-03-05					
; NUMBER OF SEQ ID NOS: 455					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 163					
; LENGTH: 2850					
; TYPE: DNA					
; ORGANISM: Homo sapiens					
US-10-205-823-163					
Query Match					
Best Local Similarity 7.9%; Score 192.6; DB 9; Length 2850;					
Matches 447; Conservative 0; Mismatches 334; Indels 24; Gaps 3;					
QY	1004	CTCAAGGGTCAAATGATCTACTTACCTCAGCAGATAGCATCTTTTCTATGTTCACCA	1063		
DB	1481	CTCAAGGCCAAATGATCTACATTGTGAATCCAGTGCACTGTGTTGGGGTCAACC	1540		
QY	1064	AGTGTCATGAACCTGCAGATTTGACAAGAGAGCGGTGTATTAAGTGACATCCCTCTG	1123		
DB	1541	TGTGTGCACAGATTAGAAGATTTTACAGGACGAGGGCTCTACCTCTCAGACATCCCAAT	1600		
QY	1124	CATGATGCCACGCGCATCTGTCTTTGGGAGAACATTTAGAGAGGAATAACAATC	1183		
DB	1601	CACAATGCATGAGGATGTGGTCTTAATAGGGGAACAAGCCGAGCTCAAGATGGCCTG	1660		
QY	1184	ACCCAAGAACTGGAATCTCTACTGACAGGCTACAGCTCACGTTAAGAGCCCTGGAAGAT	1243		
DB	1661	AAGAAGAGGCTGGGAAGCTGGAAGCTACCTTTGAGCAAGCCCACCAAGCCCTGGAGGAG	1720		
QY	1244	GAAAAGAAAAGACAGACACATTCGTCGTATCTCTCCCTCCGCTGTGTGGCAATGAG	1303		
DB	1721	GAGAAGAAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAG	1780		
QY	1304	CTGGCGCAAGCGTCCAGTGCCTGCCAAGAGATATGACAACTGTGACCATCCCTTTAGT	1363		
DB	1781	CTGTGGCAAGGGCAAGTTGTCAGGCAAGCAAGTTCAGTNATGTCACCATGCTCTCTCTCA	1840		
QY	1364	GGCATGTGGGCTCAATGCTTTCTGTAGCAAGATGCATCTGGGAAGAGCCATGAAG	1423		
DB	1841	GACATCGTTGGGTTCACTGCATCTGTGCCAGTCTC	1888		
QY	1424	ATCGTCAACCTCTCAACGACCTCTACACAGCAATTTGACACACATGACTGATTTCCCGGAAA	1483		
DB	1889	GTCAATCAACATGCTCAATGCACTGTACACTCGCTTCGACACAGCAGTGTGGAGAGCTGGA-	1947		

QY	1484	AACCCATTTGTTTTAAGGTGGAGACTGTTGGTGAACAAGTATATGACAGTGTAGTGGTTTA	1544
DB	1948	-----TGTCTACAAGGTGGAGACCATTTGGCGATGCGCTATTGTGTAGCTGGGGATTA	1999
QY	1544	CCAGAGCCATGCATTCACCATGCACCATCACTGCCACTGCCCTTGGACATGATGGNA	1603
DB	2000	CACAAGAGAGTGATCTCATCTGTTTCAGATAGGCGCTGATGGCCCTGAAGATGATGGAG	2059
QY	1604	ATTGCTGGCCAGGTT---CAAGTAGATGCTGAATCTGTTTCAGATACAAATAGGATACAC	1660
DB	2060	CTCTCTGATGAAGTTATGTCCTCCCATGGAGAACCCTATCAAGATGCGAATTTGGAATG	2119
QY	1661	ACTGGAGAGTAGTTACAGGTGTCATAGCAGCAGCGGATGCCCTCGATAGTCTCTTTTGGG	1720
DB	2120	TCTGGATCATGTTTTTGTGCTGCGTGGAGTTAAATGCCCGTTACTGTCTTTTGGGA	2179
QY	1721	AATACTGTCAACCTCACAAGCCGAACAGAAACACAGGAGAAAGGAAAAATAAATG	1780
DB	2180	ACAATGTCACCTCTGGCTAACAAATTTGAGTCTCTGCAGTGTACACGAAAAATCAATG	2239
QY	1781	TCTGAATATACATACAGATGTCTTA	1805
DB	2240	AGCCCCAACCACTTACAGATTACTCA	2264
RESULT 5			
US-10-205-823-164			
; Sequence 164, Application US/10205823			
; Publication No. US20030108963A1			
; GENERAL INFORMATION:			
; APPLICANT: Schlegel, Robert			
; APPLICANT: Monahan, John E.			
; APPLICANT: Endege, Wilson O.			
; APPLICANT: Gannavarapu, Manjula			
; APPLICANT: Gorbacheva, Bella			
; APPLICANT: Hoersch, Sebastian			
; APPLICANT: Kamatkar, Shubhangi			
; APPLICANT: Wonsey, Angela M.			
; APPLICANT: Glatt, Karen			
; APPLICANT: Zhao, Xumei			
; APPLICANT: Anderson, Dustin			
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND			
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND			
; FILE REFERENCE: MRI-044			
; CURRENT APPLICATION NUMBER: US/10/205,823			
; CURRENT FILING DATE: 2002-07-25			
; PRIOR APPLICATION NUMBER: 60/307,982			
; PRIOR FILING DATE: 2001-07-25			
; PRIOR APPLICATION NUMBER: 60/314,356			
; PRIOR FILING DATE: 2001-08-22			
; PRIOR APPLICATION NUMBER: 60/325,020			
; PRIOR FILING DATE: 2001-09-25			
; PRIOR APPLICATION NUMBER: 60/341,746			
; PRIOR FILING DATE: 2001-12-12			
; PRIOR APPLICATION NUMBER: 60/362,158			
; PRIOR FILING DATE: 2002-03-05			
; NUMBER OF SEQ ID NOS: 455			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 163			
; LENGTH: 2850			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-205-823-163			
Query Match			
Best Local Similarity 7.9%; Score 192.6; DB 9; Length 2850;			
Matches 447; Conservative 0; Mismatches 334; Indels 24; Gaps 3;			
QY	1004	CTCAAGGGTCAAATGATCTACTTACCTCAGCAGATAGCATCTTTTCTATGTTCACCA	1063
DB	1481	CTCAAGGCCAAATGATCTACATTGTGAATCCAGTGCACTGTGTTGGGGTCAACC	1540
QY	1064	AGTGTCATGAACCTGCAGATTTGACAAGAGAGCGGTGTATTAAGTGACATCCCTCTG	1123
DB	1541	TGTGTGCACAGATTAGAAGATTTTACAGGACGAGGGCTCTACCTCTCAGACATCCCAAT	1600
QY	1124	CATGATGCCACGCGCATCTGTCTTTGGGAGAACATTTAGAGAGGAATAACAATC	1183
DB	1601	CACAATGCATGAGGATGTGGTCTTAAT	

QY 1064 AGTGTCATGAACCTGGACGATTTGACAAAGGAGGAGGCTGTATCTAAGTGACATCCCTCTG 1123
Db 1562 TGTGTGCACAGATTTAGAACATTTTACAGGACGAGGCTCTACCTCTCAGACATCCCAAT 1621
QY 1124 CATGATGCCACGGCATCTTGTCTTTTGGGAGAACATTTAGAGGAAATACAACTC 1183
Db 1622 CACAATGCACTGAGGATGTGTCTTAATAGGGAAACAGCCGAGCTCAAGATGSCCTG 1681
QY 1184 ACCAAGAACTGGAATCTCACTGACAGGCTACAGCTCACGTTAAGAGCCCTGGAAGAT 1243
Db 1682 AGAAGAGGCTGGGAGCTGAAGCTACCTTTGAGCAAGCCCAACCAAGCCCTGGAGGAG 1741
QY 1244 GAAAGAAAAGACAGACATTTGCTGTATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1303
Db 1742 GAGAAGAAAAGACAGTACGCTCTGTCTCCATATTTCCCTGTGAGGTTGCTCAGCAG 1801
QY 1304 CTGGGCGACAAAGGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1363
Db 1802 CTGTGGCAAGGGCAAGTTGTGCAAGCCCAAGAAAGTTTCAAGTATGTCAACCATGCTCTCTCA 1861
QY 1364 GGCATTTGTGGCTTCAATCTTCTGTAGCAAGCATGCTGCTGAGAGGAGGCTGAAG 1423
Db 1862 GACATGTTGGTCTACTGCCATCTGCTCCAGTCTC-----ACCGTGCAG 1909
QY 1424 ATGTCACACCTCTCAACGACCTTACACAGATTTGACACACTGACTGATTTCCCGGAAA 1483
Db 1910 GTCATCACCATGCTCAATGCACTGTACACTCGCTCGACGAGGCTGAGAGGCTGGA- 1968
QY 1484 AACCCATTTGTTAAGGTGAGACTGTTGTGACAAAGTATGACAGTGTGAGTGGTTTA 1543
Db 1969 -----TGTCTAAGGTGGAGACCATTTGGCGATGCTTATTTGTAGCTGGGGGATTA 2020
QY 1544 CCAGAGCCATGCACTCACCAGTACGATCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1603
Db 2021 CACAAGAGAGTGATCTATCTGCTTACAGATAGCGCTGATGCGCTGAGATGAGGAG 2080
QY 1604 ATGCTGGCGAGGTT---CAAGTAGATGTGATCTGTTGACATACAAATAGGATACAC 1660
Db 2081 CTCTGTGATGAAGTTATGTCTCCCATGGAGAACCTATCAAGATGCGAATTTGAGCTGCA 2140
QY 1661 ACTGAGAGGTAGTTACAGGTGTCATAGGACGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1720
Db 2141 TCTGATCAAGTTTGTCTGGCTGCTGAGGTAAAGTCCCGCTTACTGCTCTTTTGA 2200
QY 1721 AATAGTCAACCTCACAAGCCGACACAAACACAGGAGAAAGGAAAAAATAATGTG 1780
Db 2201 AACAATGTCACTCTGGCTAACAAATTTGAGTCTGCGAGTGCTACCGAGAAATCAATGTC 2260
QY 1781 TCTGATATACATACAGATGCTTA 1805
Db 2261 AGCCCAACAACTTACAGATTAATCA 2285

RESULT 6

US-10-205-823-161
; Sequence 161, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wansley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044

; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 3017
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-205-823-161

Query Match 7.9%; Score 192.6; DB 9; Length 3017;
Best Local Similarity 55.5%; Pred. No. 2.3e-40;
Matches 447; Conservative 0; Mismatches 334; Indels 24; Gaps 3;

QY 1004 CTCAGGGTCAATGATCTACTTACCTGAAGCAGATAGCATACTTTTTCTATGTTCCACCA 1063
Db 1615 CTCAAAGGCCAAATGATCTACATTTTGAATCCAGTGCATCTGTTTTGGGTCAACC 1674
QY 1064 AGTGTCTGAACCTGGAGATTTGACAGGAGGCTGTATCTAAGTGCATCTCTGTC 1123
Db 1675 TGTGTGGACAGATTTAGAAATTTTACAGGACGAGGCTCTACCTCTCAGACATCCCAAT 1734
QY 1124 CATGATGCCACGCGCATCTTCTCTTTTGGGAGAACATTTAGAGGAGAAATACAACTC 1183
Db 1735 CACATGCACTGAGGATGTGTCTTAATAGGGAACAGCCGAGCTCAAGATGCCCTG 1794
QY 1184 ACCAAGAACTGGAATCTCACTGACAGGCTTACAGCTCACCTTAAAGCCCTGGAGAT 1243
Db 1795 AAGAAGAGGCTGGGAGAGCTGAAGCTACCTTTGACAGAGCCCAACCAAGCCCTGGAGGAG 1854
QY 1244 GAAAGAAAAGACAGACATTTGCTGTATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1303
Db 1855 GAGAAGAAAAGACAGATGAGCTTCTGTGCTCCATTTTCTCTGTGAGGTGCTCAGCAG 1914
QY 1304 CTGCGGCACAAAGCTCCAGTGCCTGCCAAAAGATATGACAAATGTGACCATCTCTTTAGT 1363
Db 1915 CTGTGCGAAGGCAAGTTGTGCAAGCAAGATTCAGTAATGTCACTGCTCTCTCA 1974
QY 1364 GGCATTTGGGCTTCAATGCTTTCTGTAGCAAGCATGCTATGAGAGAGGAGCCATGAAG 1423
Db 1975 GACATCGTTGGGTTCACTGCCATCTGCTCCAGTGCTC-----ACCGTGCAG 2022
QY 1424 ATCGTCAACCTCTCAAGGACCTTACACAGATTTGACACACTGACTGATTCGCGAAA 1483
Db 2023 GTCATCACCATGCTCAATGCACTGTACACTCTGCTGACACAGCATGTGTGAGAGCTGGA- 2081
QY 1484 AACCCATTTGTTTATAAGTGGAGACTGTTGGTGTGACAAAGTATATGACAGTGTGTTTA 1543
Db 2082 -----TGTCTACAAGTGGAGACCATTTGGCGATGCTATTTGTAGCTGGGGATTA 2133
QY 1544 CCAGAGCCATGCAATTCACCATGCAAGTCCATCTGCCACCTGGCTTGGACATGATGGA 1603
Db 2134 CACAAGAGAGTGTACTACTGTTTCCAGATGAGCTGTGAGGCTGAGATGATGAGGAG 2193
QY 1604 ATTGCTGGCCAGGTT---CAAGTAGATGTTGATGTTTCAAGTAAACATACATAGGATAC 1660
Db 2194 CTCTGTGATGAAGTTATGTTCTCCCATGGAGAACCTATCAAGATGCGAATTTGAGTGCAC 2253
QY 1661 ACTGAGAGGATTTACAGGTTGTATGACAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1720
Db 2254 TCTGGATCAGTTTTTGTCTGGCGCTGTTGGAGTTAAATGCCCGTCTACTGCTCTTTTGA 2313
QY 1721 AATAGTCAACCTCACAAGCCGAGAACACAGAACCCAGGAGAAAGGAAAAAATAATGTG 1780

Db 2314 AACAATGACCTCTGGCTAACAATTTGAGTCTCGAGTGTACCAGGAAATCAATGTC 2373
QY 1781 TCTGAATATACATACAGATGCTTA 1805
Db 2374 AGCCCAACAACCTTACAGATTACTCA 2398

RESULT 7
US-09-952-213D-1
; Sequence 1, Application US/09952213D
; Publication No. US20030096240A1
; GENERAL INFORMATION:
; APPLICANT: MURAD, FERID
; APPLICANT: SHARINA, IRAIDA G.
; APPLICANT: KRUMENACKER, J. S.
; APPLICANT: MARTIN, E.
; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SGC
; FILE REFERENCE: US/09/952,213D
; CURRENT APPLICATION NUMBER: 2002-08-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2430
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (286)..(2361)
US-09-952-213D-1

Query Match 7.7%; Score 187.2; DB 9; Length 2430;
Best Local Similarity 55.1%; Pred. No. 5.3e-39;
Matches 444; Conservative 0; Mismatches 338; Indels 24; Gaps 3;

QY 1003 TCTCAAGGCTCAATGATCTACTTACCTCAACACATAGCATACATCTTTCTATGTCACC 1062
Db 1380 TCTCAAGGCTCAATGATCTACTTACCTCAACACATAGCATACATCTTTCTATGTCACC 1439
QY 1063 AAGTGTCAACCTGGAGGATTTGACAGGAGAGGGCTGTATCTAAGTGACATCCCTCT 1122
Db 1440 ATGTGTGACAGCTGGAGATTTACAGGAGGGGGCTCTATCTGCGACATCCCAT 1499
QY 1123 GCATGATGCCAGCGGATCTTGTCTTTTGGGAGAACAAATTTAGAGAGGATACAACT 1182
Db 1500 TCATAACGCCCTGAGGAGTGTCTTGTATAGGGAGGACGAGCGGGCTCAAGATGGCCT 1559
QY 1183 CACCAAGAATCTGGAATCTCTACTGACAGGCTACAGCTCAGCTTAAGAGCCCTGGA 1242
Db 1560 CAGAAGAGGTGGGAGCTGAAGCAACCTGGAGCATGCCCAAGCCCTGGAGGA 1619
QY 1243 TGAAGAAGAAAGACAGACATCTGCTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1302
Db 1620 AGAAGAAGAGAGGACAGTGTGATCTGCTGCTCTATCTTCCCTCTGAGGTTGCTCAG 1679
QY 1303 GCTGGGACACAGGCTCCAGTGCCTGCCAAAGATATGACATGTGACCATCTCTTTAG 1362
Db 1680 GCTGTGGCAAGACAAATTTGTCAAGCCCAAGAAATTCAGGAGGTGACCATGCTTTCTC 1739
QY 1363 TGGCATCTGGGCTCAATGCTTCTGTAGCAAGCATGATCTGGAGAGGAGCCATGAA 1422
Db 1740 AGATATCTGGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1787
QY 1423 GATGCTCAACCTCCCAAGGCTCTTACACAGATTTGACACACTGACTGATTCGCGGAA 1482
Db 1788 GGTCAATCAGATGCTCAAGGCTCTTACACTGCTTTGACAGAGTGTGGAGAGCTGGA 1847
QY 1483 AAACCCATTTGTTAAGGTGGAGACTGTTGGTGACAAAGATATGACAGTGTGAGTGT 1542
Db 1848 -----TGCTACAGGTGGAGACCATCGGGATGATATTTGTGTGAGGTGAT 1898
QY 1543 ACCAGAGCCATGCATTCACCATGCAGGATCCATCTGCCACCTGGCCTTGACATGATGA 1602

Db 1899 GCACAGAGAGAGTGACACCCATGCTGTCCAGATAGCACTGATGGCCCTGAAGATGATGA 1958
QY 1603 AATTGTGCGCCAGG---TTCAAGTAGATGCTGTAATCTGTTTCAGATAACAATAGGGATACA 1659
Db 1959 GCTCTCCAATGAGTGCATGCTCTCCACGAGAGACCTATCAAGATCGGAATGGACTACA 2018
QY 1660 CACTGGAGAGGTAGTTTACAGGTGTATAGGACAGCGGATGCCCTCGATACTGTCTTTTGG 1719
Db 2019 TTCTGGATCAGTGTGCTGGAGTGTCTCGAGTGAAGATGCCCGGCTATTGCTGTTTGG 2078
QY 1720 GAATCTGTCAACCTCAACGCGGACAGAACACCAACAGGAGAGAAAGGAAATAATGT 1779
Db 2079 AAACAATGTCACTGTGGCTTAACAAATTTGAATCTCGCTGTGCTCGGAAATCAATGT 2138
QY 1780 GTCTGAATATACATACAGATGCTTA 1805
Db 2139 CAGCCCCACACATACAGGTTACTCA 2164

RESULT 8
US-09-917-800A-1569
; Sequence 1569, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1569
; LENGTH: 2335
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_012770
US-09-917-800A-1569

Query Match 6.0%; Score 147.2; DB 10; Length 2335;
Best Local Similarity 51.0%; Pred. No. 2.4e-28;
Matches 517; Conservative 0; Mismatches 443; Indels 54; Gaps 5;

QY 725 TTCTCAAGAGCTTTTCTTTCATATAATTTACCGGACCTAGTGTGCTACTGCTGT 784
Db 771 TTCTGTGATGCTTTTCTTTCATATAATTTACCGGACCTAGTGTGCTACTGCTGT 830
QY 785 GGCATGCTATATACAGAGATTTCTCCCGGAGCTCCAGCTTTCAGCTTCTGCTCT 844

Db 831 GGAGTGAATATTCAGAAATATGTCCTCGGAATCTTAACCCAGAAAGTTTGACATAGATGAG 890
QY 845 GTCTTCCTCGTGGTCTCGTCTCATATATGATATATAGTTTCCATGGATCTTTCTCACATC 904
Db 891 TATTTTCCATCATCCACCTCAAGTTACTTTTCAACATCTCCAGCATCTGCAAGTTTCATT 950
QY 905 AATACATGTTTTTGTATTTAGAACCAAGGAAGGATTTGTTGATGTGGAGAAATTAGAATGT 964
Db 951 AACAGTCAGTTTGTCTTGAAGACAAAGAAAGAAATGATGCC-----AAAGC 997
QY 965 GAGGATGAATGCTGCTGGGACTGAGATCAGCTGCTTACGTCTCAAGGTCAGAAATGCTAC 1024
Db 998 AAGAAGAGCCACCGATGCTCAAA-----CTCCGGGTGATGATGTTG 1043
QY 1025 TTACCTGAAGCAGATAGCATACTTTTTCTATGTTTCCAAAGTGTCTATGAACCTGGACGAT 1084
Db 1044 ATGGAGTCTCTGAGGTGATGATCTTCAATGTTTCCCAACGCTCCGACGCTGCAAGAG 1103
QY 1085 TTGACAAGGAGAGGCTGTATCTAAGTGACATCCCTCTGATGATGCGACGCGGATCTT 1144
Db 1104 CTGGAAGAGAGCAAGATGATCTTTCTGATATCGCTCCGACGACGACGACGAGGATCTC 1163
QY 1145 GTTCTTTTGGGAAACAATTTAGAGAGGAATAACAATCAACCAAGAACTGGAAATCCCTC 1204
Db 1164 ATCTCTCAACAGCAGAGGCTGGCAGAGATGGAGCTGTCTGCCAATCGGAAAGAAG 1223
QY 1205 ACTGACAGGCTACAGCTCAGTTTAAAGAGCCCTGGAAGATGAAAGAAAGAACAGACACA 1264
Db 1224 AAGGAGAGTGGTGTCTTCCATCACTGACCTGCGCATCGAAGAAAGAACAGAGAC 1283
QY 1265 TTGCTGTATCT 1324
Db 1284 TTGCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1343
QY 1325 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
Db 1344 GCTGAGGAGATTTGAACATGTAACATCTTCAATCTTTCAGCGATGTTGTGACATTTACCA 1403
QY 1385 TTCTGTAGCAAGCATGATCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1444
Db 1404 ATCTGTCACGCTGTGAACCT-----ATCCAAATCTGAACATGCTGAATCA 1451
QY 1445 CTCTACACAGATTTACACACTGACTGATCTCCGGAAGAAACCCATTTGTTTATAGGTG 1504
Db 1452 ATGACTCCAAAGTTTACAGGTTTAAACAGTGTCTCATGA-----TGCTCAAAAGTA 1502
QY 1505 GAGACTGTTGGTGACAAGTATATGACAGTGAGTGGTTTACCAAGGAGGAGGAGGAGGAGGAG 1564
Db 1503 GAACAATAGGGATGCTTACATGTTGGTGGTGGAGTACCACTACCCGTTGAAAGCCAT 1562
QY 1565 GCAGATCCATCTGCCACCTGGCCTTGGACATGA-----TGGAAATGCTGCCAGGTT 1618
Db 1563 GCTCAAGAGTCCGCAATTTGCTCTGGGATGAGAATTTCTCAAAAGAGTGAAT 1622
QY 1619 CAAGTAGATGCTGAATCTGTTCAAGATAACATAGGAGTACACACTGGAGAGGTAGTTACA 1678
Db 1623 CTGTCACGTGGGAAACCTATCCAGATCAGAGTGGGAATCCACACTGGACCAAGTCTTAGCA 1682
QY 1679 GGTGTATAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1732
Db 1683 GGTGTTGGGACAGAGATGCTCTGGTACTGCTGTTGTTGGTGCACACTGTAAAC 1736

RESULT 9

US-09-952-213D-4

; Sequence 4, Application US/09952213D
; Publication No. US20030096240A1

; GENERAL INFORMATION:

; APPLICANT: MURAD, FERID
; APPLICANT: SHARINA, IRAIDA G.
; APPLICANT: KRUMENACKER, J. S.
; APPLICANT: MARTIN, E.

; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SGC

; FILE REFERENCE: UTSH:252US
; CURRENT APPLICATION NUMBER: US/09/952,213D
; CURRENT FILING DATE: 2002-08-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4

; LENGTH: 7697

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: modified_base

; LOCATION: (605)..(6955)

; OTHER INFORMATION: N = A, C, T/U OR G

US-09-952-213D-4

Query Match 5.4%; Score 131.8; DB 9; Length 7697;

Best Local Similarity 56.4%; Pred.No. 7.3e-24;
Matches 286; Conservative 3; Mismatches 205; Indels 13; Gaps 2;

QY 1003 TCTCAAGGGTCAAAATGATCTACTTACCTGAAGCAGATAGCATACTTTTCTATGTTCCACC 1062
Db 4262 TCTCAAAAGGTCAAAATGATCTACTTACCTGAAGCAGATAGCATACTTTTCTATGTTCCACC 4321
QY 1063 AAGTGTATCAACCTCGACGATTTGACAAGGAGGAGGCTGTATCTAAGTGACATCCCTCT 1122
Db 4322 ATGTGTGGACAGCTGGAAGATTTCAAGGACGGGGCTCTATCTGTCCGACATCCCAAT 4381
QY 1123 GCATGATGTCACCGCGGATCTTGT-TCTTTTGGAGAACAAATTTAGAGAGGAATACAAAC 1181
Db 4382 TCATAACGCCCTGAGGATGTTTTCATGATAGGGGAGAGGACGCGGCTCAAGATGGCC 4441
QY 1182 TCACCAAGAACTGGAATCTCTCACTGACAGGCTACAGCTACGTTAAGAGCCCTCGAAG 1241
Db 4442 TCAAGAAGAGGTGGGGAAGCTGAAGCAACCTGGAGCATGCCACCAAGCCCTGGAGG 4501
QY 1242 ATGAAAAGAAAAGACAGACACATCTGCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1301
Db 4502 ARGAGAAGAGGAGGACAGATGATGCTGCTGCTCTATCTCTCTCTCTCTCTCTCTCTCT 4561
QY 1302 AGTCTGGGACACAGCTCCAGTCCCAAGATATGACAAATGTGACCATCTCTTTA 1361
Db 4562 AGCTGTGGCAAGACAAATTTGCAAGCCCAAGAAATTCAGGAGGTCCACCATCTCTTCT 4621
QY 1362 GTGGCATGTTGGGCTTCAATGCTTCTGTAGCAAGCATGCTGTGAGAGAGGAGCCATGA 1421
Db 4622 CAGATATGTAGGGTTCACTGCTATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4669
QY 1422 AGATCTCAACCTCTCAAGGACCTCTACACAGATTTGACACACTGACTGATTCGGGA 1481
Db 4670 AGGTATCATCGATGCTCAACGCTCTCTACACTCGCTTTGACAGGAGGTGGAGGTGG 4729
QY 1482 AAAAAACCAATTTGTTTATAAGGTGGAGA 1508
Db 4730 ATGCTCAAGGTAGGAGGAGGTGAAA 4756

RESULT 10

US-09-952-213D-6/c

; Sequence 6, Application US/09952213D

; Publication No. US20030096240A1

; GENERAL INFORMATION:

; APPLICANT: MURAD, FERID

; APPLICANT: SHARINA, IRAIDA G.

; APPLICANT: KRUMENACKER, J. S.

; APPLICANT: MARTIN, E.

; TITLE OF INVENTION: GENOMIC ORGANIZATION OF MOUSE AND HUMAN SGC

; FILE REFERENCE: UTSH:252US

; CURRENT APPLICATION NUMBER: US/09/952,213D

; CURRENT FILING DATE: 2002-08-16

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 177556

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 10:46:42 ; Search time 5166.6 Seconds
(without alignments)
16983.119 Million cell updates/sec

Title: US-09-762-767A-1
Perfect score: 3015
Sequence: 1 cccctatgcgattggcg.....aaaaaaaaaaaaaaaa 3015

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues 4109280
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vt.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3015	100.0	3015	6	AX024599	AX024599 Sequence
2	3015	100.0	3015	9	HSRNASGC	Y15723 Homo sapien
3	2855.4	94.7	3004	9	HSGCSRA	X66534 H.sapiens s
4	2378.2	78.9	2857	9	BC028384	BC028384 Homo sapi
5	2312.6	76.7	2827	9	AK096588	AK096588 Homo sapi
6	2258	74.9	2258	9	HSU58855	U58855 Homo sapien
7	1620.6	53.8	3261	4	BTGUCY73	X54014 Bovine mRNa
8	1573.6	52.2	4775	10	RNU60835	U60835 Rattus norv
9	1552	51.5	2445	10	AF297082	AF297082 Mus muscu
10	1536.8	51.0	2668	10	RATGCSB	M57405 Rat solubie
11	716.2	23.8	1571	9	AF233746S1	AF233746 Homo sapi
12	714.6	23.7	17556	2	AC021433	AC021433 Homo sapi
13	714.6	23.7	196537	9	AC104083	AC104083 Homo sapi
14	660.2	21.9	3022	5	AB062171	AB062171 Takifugu
15	659.4	21.9	2570	5	AB000849	AB000849 Oryzias l
16	548	18.2	2854	6	AX281673	AX281673 Sequence
17	548	18.2	2854	9	HSGCA2	X63282 H.Sapiens m
18	547.4	18.2	2657	10	AF109963	AF109963 Rattus no
19	536.2	17.8	718	9	AF233746S6	AF233750 Homo sapi
20	498	16.5	155745	2	AC098267	AC098267 Rattus no
21	498	16.5	218756	2	AC128391	AC128391 Rattus no
22	494.2	16.4	299480	2	AC121563	AC121563 Mus muscu
23	493.2	16.4	934	9	AF233746S2	AF233747 Homo sapi
24	446.6	14.8	3047	9	HSRNASCA2	Z50053 H.sapiens m
25	329.4	10.9	5809	3	AF062750	AF062750 Manduca s
26	317.2	10.5	2345	3	DMU27117	U27117 Drosophila
27	314.8	10.4	37274	2	AC012920	AC012920 Drosophila
28	314.8	10.4	178791	3	AC008218	AC008218 Drosophila
29	314.8	10.4	202929	3	AC007892	AC007892 Drosophila
30	314.8	10.4	224100	3	AE003770	AE003770 Drosophila
31	314	10.4	2385	3	AY060654	AY060654 Drosophila
32	301.2	10.0	2328	3	S57126	S57126 dgel-guanyl
33	297.8	9.9	16504	5	AB022280	AB022280 Oryzias l
34	291	9.7	10941	5	AB062169	AB062169 Takifugu
35	288.8	9.6	69135	2	AC102100	AC102100 Mus muscu
36	281.2	9.3	206456	2	AC112402	AC112402 Rattus no
37	277.4	9.2	165258	9	AP001282	AP001282 Homo sapi
38	277.4	9.2	168732	2	AC024050	AC024050 Homo sapi
39	276.8	9.2	661	10	AB079780	AB079780 Rattus no
40	276.2	9.2	62894	2	AC113467	AC113467 Mus muscu
41	275	9.1	328	11	G19426	G19426 human STS S
42	275	9.1	188237	2	AP000906	AP000906 Homo sapi
43	263.4	8.7	155397	2	AC128853	AC128853 Rattus no
44	242	8.0	143052	2	AC079207	AC079207 Homo sapi
45	192.6	6.4	2443	6	AX024601	AX024601 Sequence

ALIGNMENTS

RESULT 1
AX024599
LOCUS AX024599
DEFINITION Sequence 1 from Patent DE19837015.
ACCESSION AX024599
VERSION AX024599.1 GI:10184739
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3015)
AUTHORS Poller.W., Schmidt,H. and Zabel,U.
TITLE New human soluble guanylate cyclase alpha/beta1 and the nucleic acid encoding the subunits, useful for producing diagnostic

antibodies, and for somatic gene therapy of arteriosclerosis	
JOURNAL	PATENT: DE 19837015-A 1 24-FEB-2000;
FEATURES	Location/Qualifiers
source	1..3015
BASE COUNT	877 a 680 c 716 g 742 t
ORIGIN	
Query Match	
Best Local Similarity	100.0%; Score 3015; DB 6; Length 3015;
Matches 3015; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 CCCTTATGGGATGGCGGCTGCAGAGACGAGGACTCAGTTCCTCCCTGCGCTAGTCTGAG 60
Db	1 CCCTTATGGGATGGCGGCTGCAGAGACGAGGACTCAGTTCCTCCCTGCGCTAGTCTGAG 60
Qy	61 CCTAGTGGTGGGACTCAGTCTCAGATCAGTTCAGTTCAGAGAGAGGTTTTCAGTGCAGAGTT 120
Db	61 CCTAGTGGTGGGACTCAGTCTCAGATCAGTTCAGTTCAGAGAGAGGTTTTCAGTGCAGAGTT 120
Qy	121 TTCCTACACTTTTCTCGCTAGAGCAGCGAGCAGCTTGAACAGACCCAGGCGGAGGAC 180
Db	121 TTCCTACACTTTTCTCGCTAGAGCAGCGAGCAGCTTGAACAGACCCAGGCGGAGGAC 180
Qy	181 ACCTGTGGGAGGAGCGCTGGAGGAGCTTAGAGACCCGCGGCGTGTATCTCACC 240
Db	181 ACCTGTGGGAGGAGCGCTGGAGGAGCTTAGAGACCCGCGGCGTGTATCTCACC 240
Qy	241 ATGTGGGATTTGGAGGCGGCGCTGGAGTGTCTAGAGATCCGGAAGACAGCCCGGAG 300
Db	241 ATGTGGGATTTGGAGGCGGCGCTGGAGTGTCTAGAGATCCGGAAGACAGCCCGGAG 300
Qy	301 GTGTGCAAGCCACCAAGACTGCGGCTTGGAGAAAGCGTGGAGGAGCGTGAGCGGGGCGACCGCG 360
Db	301 GTGTGCAAGCCACCAAGACTGCGGCTTGGAGAAAGCGTGGAGGAGCGTGAGCGGGGCGACCGCG 360
Qy	361 TCTCCGCGCTGTCTGACCCCTGFCGCTGAGCTGCTGAGTGCATGACATGACATCCAG 420
Db	361 TCTCCGCGCTGTCTGACCCCTGFCGCTGAGCTGCTGAGTGCATGACATGACATCCAG 420
Qy	421 TTACCACTGCTTGAATTCATAGTGGCTTCTGTTCTGCTGCTCATATAGACTACAG 480
Db	421 TTACCACTGCTTGAATTCATAGTGGCTTCTGTTCTGCTGCTCATATAGACTACAG 480
Qy	481 CTCATCAGGAGGATCGCAGCGGTAAGAGACACCAACACCATGTTCTGCAGAGCT 540
Db	481 CTCATCAGGAGGATCGCAGCGGTAAGAGACACCAACACCATGTTCTGCAGAGCT 540
Qy	541 CAAGGATCTCAAGTACAGAGAGTGTCTTCTTCTTACTGGCAGCAGGTCAGAGTTC 600
Db	541 CAAGGATCTCAAGTACAGAGAGTGTCTTCTTCTTACTGGCAGCAGGTCAGAGTTC 600
Qy	601 TAACGAGCTTTCAGAGGAGCAGCAGGAAGCTCAGAGAGCTGCAAGCAACCGTGCCCAT 660
Db	601 TAACGAGCTTTCAGAGGAGCAGCAGGAAGCTCAGAGAGCTGCAAGCAACCGTGCCCAT 660
Qy	661 CTGTCAAGACATCTCTCAGAGAACATACAGAAAGTCTTCTTCTTACTGGCAGCAGGTCAG 720
Db	661 CTGTCAAGACATCTCTCAGAGAACATACAGAAAGTCTTCTTCTTACTGGCAGCAGGTCAG 720
Qy	721 GAGCGGAGTCTATCTTACACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCAGAGTT 780
Db	721 GAGCGGAGTCTATCTTACACTTTGGCAGAGAGTATTTGCAAACTGATTTTCCAGAGTT 780
Qy	781 TGAAGGCTGAATGTTTTCACCTTCAGAGAACATTTGGCAAGCACAAAATAAAGAAAGCAG 840
Db	781 TGAAGGCTGAATGTTTTCACCTTCAGAGAACATTTGGCAAGCACAAAATAAAGAAAGCAG 840
Qy	841 GAAATCTTTGGAAGAGAGACTTTGAAAGAACTTTGAGAGCAAGCAGTTCGACGAG 900
Db	841 GAAATCTTTGGAAGAGAGACTTTGAAAGAACTTTGAGAGCAAGCAGTTCGACGAG 900

Qy	901 AGTTCAGTGGAGGTTATCAAGAATCTCTTGGTGAAGAGGTTTTTAAATATGTTACGA 960
Db	901 AGTTCAGTGGAGGTTATCAAGAATCTCTTGGTGAAGAGGTTTTTAAATATGTTACGA 960
Qy	961 GGAAGATGAAACATCCTTGGGGTGGTGGAGCACCTTAAAGATTTTTTAAACAGCTT 1020
Db	961 GGAAGATGAAACATCCTTGGGGTGGTGGAGCACCTTAAAGATTTTTTAAACAGCTT 1020
Qy	1021 CAGTACCTTCTGAACACAGAGCAGCCATTGCCAGAGAGGAGGAGGAGGAGGAGGAGGCTGA 1080
Db	1021 CAGTACCTTCTGAACACAGAGCAGCCATTGCCAGAGAGGAGGAGGAGGAGGAGGCTGA 1080
Qy	1081 GGAGCGCTCCATCTTATGCGCTGGATAGAGGAGGATTTTCTACATCTTTTACTACTTCTT 1140
Db	1081 GGAGCGCTCCATCTTATGCGCTGGATAGAGGAGGATTTTCTACATCTTTTACTACTTCTT 1140
Qy	1141 CCCTAAGAGAACACCTCCCTGATCTTCCCGGATCATAAAGGAGCAGCTGCTCACGTATT 1200
Db	1141 CCCTAAGAGAACACCTCCCTGATCTTCCCGGATCATAAAGGAGCAGCTGCTCACGTATT 1200
Qy	1201 ATATGAACGGAAGTGAAGTGTGTTAATGCCTCCCTGCTTCCATATGATTCACGCGA 1260
Db	1201 ATATGAACGGAAGTGAAGTGTGTTAATGCCTCCCTGCTTCCATATGATTCACGCGA 1260
Qy	1261 GTTTGTGAATCAGCCCTACTTGTGTACTCCCTGTTACATGAAAGACACCAAGCCATCCCT 1320
Db	1261 GTTTGTGAATCAGCCCTACTTGTGTACTCCCTGTTACATGAAAGACACCAAGCCATCCCT 1320
Qy	1321 GTCCCCAGAAACCCAGTCCCTGCTGGTGGATTTCCACATCGCTATTTTGCAGACATT 1380
Db	1321 GTCCCCAGAAACCCAGTCCCTGCTGGTGGATTTCCACATCGCTATTTTGCAGACATT 1380
Qy	1381 TCCATTCCATTTTCATGTTTGCACAAAGATATGACAAATTTCTCAATTTTGGCAATGSCATCAG 1440
Db	1381 TCCATTCCATTTTCATGTTTGCACAAAGATATGACAAATTTCTCAATTTTGGCAATGSCATCAG 1440
Qy	1441 AAGGCTGATGAACAGGAGAGACTTTCAAGGAAAGCCCTTAATTTTGAAGAACTTTGAAAT 1500
Db	1441 AAGGCTGATGAACAGGAGAGACTTTCAAGGAAAGCCCTTAATTTTGAAGAACTTTGAAAT 1500
Qy	1501 TCTGACTTCCAAATCAACAGAGCTTTAGCGGATCATGACTATGTTGAATGACAGTT 1560
Db	1501 TCTGACTTCCAAATCAACAGAGCTTTAGCGGATCATGACTATGTTGAATGACAGTT 1560
Qy	1561 TGTGTACGAGTGGAGAGTGGGACACTCTGTGAAGAAATCTTCAAGGCTTATGGACCT 1620
Db	1561 TGTGTACGAGTGGAGAGTGGGACACTCTGTGAAGAAATCTTCAAGGCTTATGGACCT 1620
Qy	1621 CAAAGGCCAAATGATCTACATTTGTAATCCAGTGCATCTGTTTGGGGTGCACCTG 1680
Db	1621 CAAAGGCCAAATGATCTACATTTGTAATCCAGTGCATCTGTTTGGGGTGCACCTG 1680
Qy	1681 TGTGGCAGATTTAGAGATTTTACAGCAGGAGCTTCTACCTCTCAGACATCCCAATCA 1740
Db	1681 TGTGGCAGATTTAGAGATTTTACAGCAGGAGCTTCTACCTCTCAGACATCCCAATCA 1740
Qy	1741 CAATGCACTGAGGAGTGTGCTTAAATAGGGGAACAGCCGAGCTCAAGATGSCCTGAA 1800
Db	1741 CAATGCACTGAGGAGTGTGCTTAAATAGGGGAACAGCCGAGCTCAAGATGSCCTGAA 1800
Qy	1801 GAAGAGGCTGGGGAAGCTGAGGCTACCCCTTGAAGAGCCCAAGCCCTGGAGGAGGA 1860
Db	1801 GAAGAGGCTGGGGAAGCTGAGGCTACCCCTTGAAGAGCCCAAGCCCTGGAGGAGGA 1860
Qy	1861 GAAGAAAGACAGTACAGCTTCTGCTCCATATTTCCCTGTGAGTTCCTCAGCAGCT 1920
Db	1861 GAAGAAAGACAGTACAGCTTCTGCTCCATATTTCCCTGTGAGTTCCTCAGCAGCT 1920
Qy	1921 GTGGCAAGGCAAGTTGTGCAAGCCCAAGAGTTTCAATATGTCACCATGCTCTTCTCAGA 1980
Db	1921 GTGGCAAGGCAAGTTGTGCAAGCCCAAGAGTTTCAATATGTCACCATGCTCTTCTCAGA 1980

QY 1381 TCATTCATTCATGTTTGACAAAGATATGACAAATTCGCAATTTGGCAATGGCATCAG 1440
Db 1381 TCATTCATTCATGTTTGACAAAGATATGACAAATTCGCAATTTGGCAATGGCATCAG 1440
QY 1441 AAGGCTGATGAACAGGAGAGACTTTCAGGAAGAGCCTAATTTTGAAGAATACTTTCAAAAT 1500
Db 1441 AAGGCTGATGAACAGGAGAGACTTTCAGGAAGAGCCTAATTTTGAAGAATACTTTGAAAT 1500
QY 1501 TGTGACTCCAAAATCAACAGACGTTTACGGGATCATGACTATGTTGAATATGCAGTT 1560
Db 1501 TGTGACTCCAAAATCAACAGACGTTTACGGGATCATGACTATGTTGAATATGCAGTT 1560
QY 1561 TGTGTACAGTACAGGATGAGCAACACTCTGTGAAGAAATCTTCAAGGGTTATGGACCT 1620
Db 1561 TGTGTACAGTACAGGATGAGCAACACTCTGTGAAGAAATCTTCAAGGGTTATGGACCT 1620
QY 1621 CAAAGGCCAAATGATACATGTTGAATCCAGTGCATCTTGTGTTTGGGTCACCCCTG 1680
Db 1621 CAAAGGCCAAATGATACATGTTGAATCCAGTGCATCTTGTGTTTGGGTCACCCCTG 1680
QY 1681 TGTGGACAGATTAGAAAGATTTTACAGGACGAGGGCTTACCTCTCAGACATCCCAATTCA 1740
Db 1681 TGTGGACAGATTAGAAAGATTTTACAGGACGAGGGCTTACCTCTCAGACATCCCAATTCA 1740
QY 1741 CAATGCACGTGAGGATGCTGCTTAATAGGGGAACAGCCGAGCTCAAGATGGCTGAA 1800
Db 1741 CAATGCACGTGAGGATGCTGCTTAATAGGGGAACAGCCGAGCTCAAGATGGCTGAA 1800
QY 1801 GAAGAGCTGGGGAAGCTGAAGCTACCTTGAAGCAAGCCCAAGCCCTGGAGGAGGA 1860
Db 1801 GAAGAGCTGGGGAAGCTGAAGCTACCTTGAAGCAAGCCCAAGCCCTGGAGGAGGA 1860
QY 1861 GAAGAAAAAGACAGTACGCTTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCT 1920
Db 1861 GAAGAAAAAGACAGTACGCTTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCT 1920
QY 1921 GTGGCAAGGCAAGTTGTCAAGCCCAAGAGTTCACTAATGTACCATGCTCTTCACGA 1980
Db 1921 GTGGCAAGGCAAGTTGTCAAGCCCAAGAGTTCACTAATGTACCATGCTCTTCACGA 1980
QY 1981 CATCGTTGGGTTCACTGCCATCTGCTCCAGCTGCTCAGGCTGCAGGTCATCACCATGCT 2040
Db 1981 CATCGTTGGGTTCACTGCCATCTGCTCCAGCTGCTCAGGCTGCAGGTCATCACCATGCT 2040
QY 2041 CAATGCACGTACACTCGCTTGCACAGCAGTGTGGAGAGCTGGATGTCTACAAGTGA 2100
Db 2041 CAATGCACGTACACTCGCTTGCACAGCAGTGTGGAGAGCTGGATGTCTACAAGTGA 2100
QY 2101 GACCATTTGGCGATGCTTATGTTAGTGTGGGGATTACACAAGAGAGTGTACTCATGC 2160
Db 2101 GACCATTTGGCGATGCTTATGTTAGTGTGGGGATTACACAAGAGAGTGTACTCATGC 2160
QY 2161 TGTTCAGATAGCGCTGATGCCCTGAAGATGATGGAGCTCTCTGTGAAGTTATGTCCTC 2220
Db 2161 TGTTCAGATAGCGCTGATGCCCTGAAGATGATGGAGCTCTCTGTGAAGTTATGTCCTC 2220
QY 2221 CCATGAGAACCTATCAAGATGCGAATTTGGACTGCACCTCTGGATCAGTTTTTGGCTGG 2280
Db 2221 CCATGAGAACCTATCAAGATGCGAATTTGGACTGCACCTCTGGATCAGTTTTTGGCTGG 2280
QY 2281 CGTTGGAGTTAAATSCCCGTTTACTGCTCTTTTGGAAACAATGTCTGCTGCTAACAA 2340
Db 2281 CGTTGGAGTTAAATSCCCGTTTACTGCTCTTTTGGAAACAATGTCTGCTGCTAACAA 2340
QY 2341 ATTGAGTCTGAGTGTACCCAGGAAATATGTCAGCCCAACAACTTACAGATTACT 2400
Db 2341 ATTGAGTCTGAGTGTACCCAGGAAATATGTCAGCCCAACAACTTACAGATTACT 2400
QY 2401 CAAAGACTGTCTGGTTTGTGTTTACCCCTGATCAAGGGAGAACTTCCACCAAACTT 2460
Db 2401 CAAAGACTGTCTGGTTTGTGTTTACCCCTGATCAAGGGAGAACTTCCACCAAACTT 2460
QY 2461 CCCTAGTGAATCCCGGATCTGCCATTTCTGGATGCTTACCACACAGGACAACTC 2520

Db 2461 CCTAGTGAATCCCGGATCTGCCATTTCTGGATGCTTACCACAAAGGACAACTC 2520
QY 2521 AAAACCATCTTCCAAAGAAAGATGTGAAGATGCAATGCAATTTTGTAGGCAAGC 2580
Db 2521 AAAACCATCTTCCAAAGAAAGATGTGAAGATGCAATGCAATTTTGTAGGCAAGC 2580
QY 2581 ATCAGGAATAGATTAGCAACCTATATACCTATTTATTAAGTCTTTGGGGTTTGACTCAT 2640
Db 2581 ATCAGGAATAGATTAGCAACCTATATACCTATTTATTAAGTCTTTGGGGTTTGACTCAT 2640
QY 2641 AAGATGTAGAGCCTCTGAAGACCTTTAGGATTTAGGATTTAGATGGCTTAACAAGCAGTATTA 2700
Db 2641 AAGATGTAGAGCCTCTGAAGACCTTTAGGATTTAGGATTTAGATGGCTTAACAAGCAGTATTA 2700
QY 2701 AAATTTTCAGGAGCAAGTCAATCTTCTCTCTTTTAAACATCACAAATGTACTCACCT 2760
Db 2701 AAATTTTCAGGAGCAAGTCAATCTTCTCTCTTTTAAACATCACAAATGTACTCACCT 2760
QY 2761 CAGTACTTCAGCTCTTCAAGAAAAAACCCTTAAAAAGCTACTTTTGTGGAGTA 2820
Db 2761 CAGTACTTCAGCTCTTCAAGAAAAAACCCTTAAAAAGCTACTTTTGTGGAGTA 2820
QY 2821 TTTCTATTATTAACACGACCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 2880
Db 2821 TTTCTATTATTAACACGACCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 2880
QY 2881 TCAGATAAATGTAGTCAATTTGACAACTGATGAGTCACTGCAATCTCATATCTCTGCT 2940
Db 2881 TCAGATAAATGTAGTCAATTTGACAACTGATGAGTCACTGCAATCTCATATCTCTGCT 2940
QY 2941 GGAATGCCATGTTTAAAGTGTGTTTGTGATGTTGTCGTCACAAAAAAGAAAAA 3000
Db 2941 GGAATGCCATGTTTAAAGTGTGTTTGTGATGTTGTCGTCACAAAAAAGAAAAA 3000
QY 3001 AAAAAAAGAAAAA 3015
Db 3001 AAAAAAAGAAAAA 3015

RESULT 3
HSCSAA
LOCUS
DEFINITION H.sapiens soluble guanylate cyclase large subunit mRNA.
ACCESSION X66534
VERSION X66534.1 GI:31683
KEYWORDS cytoplasmic protein; GTP pyrophosphate-lyase; Guanylate cyclase.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3004)
AUTHORS Giulii,G., Scholl,U., Bulle,F. and Guellaen,G.
TITLE Molecular cloning of the cDNAs coding for the two subunits of soluble guanylyl cyclase from human brain
JOURNAL FEBS Lett. 304 (1), 83-88 (1992)
MEDLINE 92316204
PUBMED 1352257
REFERENCE 2 (bases 1 to 3004)
AUTHORS Guellaen,G.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-1992) Georges Guellaen, Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, FRANCE
FEATURES
Location/Qualifiers
source 1. .3004
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q31.3-q33"
/tissue_type="brain, frontal lobe"
/clone_lib="lambda GT10"
523..2676
/gene="GC-S-alpha-3"

CDS	523.. 2676 /gene="GC-S-alpha-3" /EC_number="4.6.1.2" /note="soluble guanylate cyclase large subunit" /codon_start=1 /product="guanylate cyclase" /protein_id="CAA47145.1" /db_xref="gi:31684" /translation="MFCFKLKDIAITGCEPFSLLAPGVNSESSEAAAGSSSECKATV PICODIPKNIQESLPQRKTSRSRVLYHTLAESICKLIFFEPFRLNALQRTYLAHKHI KESRSLKEDEFPKTAIQAOVQQPSVELSNKLRYKYTRMKSTSGWLEAPUKI FKQLQYPSETOPILPRSRKRGOLEDASILICDELDDHLHYFFPYFRKRTSLIPLGIK AAHVLYTEVEVSLPPCFHNDCSEFNOPYLLYSVMKSTKPSLSPKSQSSLVIP TSLFCXKTEVFHFMDKDMTILOFGNGIRLLMNRRDFOGFNFYEFLTPKINOTPSG INTMLNMOFVVVRWRDNVSVKSRVDMLKGOMIYIVESSAILFLPGSPCVDRLEDFTG RGYLISLDIPIHNALRDVYLIGEORADGLKKRLKALKALEQAHQALREBKKTVDL LCISIFPEVAOLMQOVQAKRESNTMLFDIVGFETACISOCSPHQVLTIMLNALYT RDQCQGLDYKVETITAMPYLVIGLGHKSDTHAQVIALMALKNMELSDENVSPRIE PTKMRIGLHSVSVAGVYVGMKPXYCLFGNNVTLANKFESCVPKRIINVSPITYRLLK DCPGVFPTPRSEELPNPFPEIGICHFIHDAYOOGTNSKPCPKOKKDVEDASOFFRQS INRLIATVIPIYKSLGPDLSLMCRASESTLGIVDG"
	BASE COUNT 874 a 679 c 710 g 741 t
ORIGIN	
	Query Match 94.7%; Score 2855.4; DB 9; Length 3004; Best Local Similarity 98.6%; Pred. No. 0; Matches 2976; Conservative 0; Mismatches 26; Indels 15; Gaps 9;
Qy	1 CCTTATGGCGATTGGGGCGGTGCAGAGCAGAGCTAGTTCCCGTCGCCCTAGTCGTAG 60
Dd	1 CCTTATGGCGATTGGGGCGGTGCAGAGCAGAGCTAGTTCCCGTCGCCCTAGTCGTAG 60
Qy	61 CTTAGTGCGTGGACTCAGCTCAGAGTCAGTTTTTCCAGAAGCAGGTTTCAGTCAGAGTT 120
Dd	61 CTTAGTGCGTGGACTCAGCTCAGAGTCAGTTTTTCCAGAAGCAGGTTTCAGTTGAGAGTT 120
Qy	121 TTCTACACTTTTCTCGCGTAGAGCAGCAGAGCCCTGGAAACAGACCAGCGGGAGGAC 180
Dd	121 TTCTACACTTTTCTCGCGTAGAGCAGCAGAGCCCTGGAAACAGACCAGCGGGAGGAC 180
Qy	181 ACTGTGGGGAGGAGCGCTGTGAGGAGCTTAGAGACCCAGCGGGCGGTGATCTCAC 240
Dd	181 ACCTGTGGGGAGGAGCGCTGTGAGGAGCTTAGAGACCCAGCGGGCGGTGATCTCAC 240
Qy	241 ATGTGCGGATTTGCAGAGCGCGCCCTGAGCTGCTAGAGATCCGGAAGACACGCCCGGAG 300
Dd	241 ATGTGCGGATTTGCAGAGCGCGCCCTGAGCTGCTAGAGATCCGGAAGACACGCCCGGAG 300
Qy	301 GTGTGCGAAGCCCAAGACTCGCGCTCTTGGAGAAAGCGTGAGCAGGGGGCCACCGCG 360
Dd	301 GTGTGCGAAGCCCAAGACTCGCGCTCTTGGAGAAAGCGTGAGCAGGGGGCCACCGCG 360
Qy	361 TCTCCGCGCTGTGCACCTGTGCGCTGAGCTGCGCTGAGCAGTACAATGCAATCCAG 420
Dd	361 TCTCCG-GCCTGTGCACCTGTGCGCTGAGCTGCGCTGAGCAGTACAATGCAATCCAG 419
Qy	421 TTACCATGCTCTTCAATTTGATAGTGGCTTCTGTTGTGTCAGTCTCATATAAGAACTACAG 480
Dd	420 TTACCATGCTCTTGAATTTGATAGTGGCTTCTGTTGTGTCAGTCTCATATAAGAACTACAG 479
Qy	481 CTCATCAGGAGGAGATCGCAGCAGGGTAAAGACACCAACCACTGTTCTGCACGAAGCT 540
Dd	480 CTCATCAGGAGGAGATCGCAGCAGGGTAAAGACACCAACCACTGTTCTGCACGAAGCT 539
Qy	541 CAAGGATCTCAAGATCACAGGAGTGTCTTTCTCTTACTGGCACCAGGTCAAGTTC 600
Dd	540 CAAGGATCTCAAGATCACAGGAGTGTCTTTCTCTTACTGGCACCAGGTCAAGTTC 599
Qy	601 TAACAGTCTTTCAGAGGAGGAGCAGGAGCTCAGAGAGCTGCAAGACCAACCGTGGCCAT 660
Dd	600 TAACAGTCTTTCAGAGGAGGAGCAGGAGCTCAGAGAGCTGCAAGACCAACCGTGGCCAT 659

```
Db 1731 CAATGCATGAGGATGGTCTTAATAGGGAACAAGCCGAGCTCAAGATGGCTGAA 1790
QY 1801 GAAGAGGCTGGGAAAGCTGAAGGCTTACCTTTGAGCAAGCCACCAGCCCTGGAGGGA 1860
Db 1791 GAAGAGGCTGGGAAAGCTGAAGGCTTACCTTTGAGCAAGCCACCAGCCCTGGAGGGA 1850
QY 1861 GAAGAAAAGACAGTAGACCTTCTGTGCTCATAATTTCCCTGTGAGGTTGCTCAGCAGCT 1920
Db 1851 GAAGAAAAGACAGTAGACCTTCTGTGCTCATAATTTCCCTGTGAGGTTGCTCAGCAGCT 1910
QY 1921 GTGGCAAGGCAAGTTGCAAGCAAGCAAGTTCAAGTAATGTCACCATGCTCTTCTCAGA 1980
Db 1911 GTGGCAAGGCAAGTTGCAAGCAAGCAAGTTCAAGTAATGTCACCATGCTCTTCTCAGA 1970
QY 1981 CATCGTTGGGTTCACTGCGCATCTGCTCCAGTGTCTCAGGCTGCAGGTCATCACCATGCT 2040
Db 1971 CATCGTTGGGTTCACTGCGCATCTGCTCCAGTGTCTCAGGCTGCAGGTCATCACCATGCT 2030
QY 2041 CAATGCATGTACACTCGCTCGACCAAGAGTGTGGAGAGTGTGAGTGTCAAGAGTGA 2100
Db 2031 CAATGCATGTACACTCGCTCGACCAAGAGTGTGGAGAGTGTGAGTGTCAAGAGTGA 2090
QY 2101 GACCATTCGCGATGCTATTGTGTAGC-TGGGGGATTACAAAGAGAGTGTATCTCATG 2159
Db 2091 GACCATT-GCGATGCTATTGTGTGGCTTGGGGGATTACAAAGAGAGTGTATCTCATG 2149
QY 2160 CTGTTTCAGATAGCGTGTGCGCTGAAGATGATGGAGTCTCTCATGAAGTTATGTCCTC 2219
Db 2150 CTGTTTCAGATAGCGTGTGCGCTGAAGATGATGGAGTCTCTCATGAAGTTATGTCCTC 2209
QY 2220 CCCATGGAGAACCTATCAAGATGCGAATTTGAGTGTGAGTGTGAGTGTGAGTGTGAGG 2279
Db 2210 CCCATGGAGAACCTATCAAGATGCGAATTTGAGTGTGAGTGTGAGTGTGAGTGTGAGG 2269
QY 2280 TCGTTGGAGTAAATCCCGTACTGCTCTTTTGGAAACAATGCTACTCTGGGTAACA 2339
Db 2270 TCGTTGGAGTAAATCCCGTACTGCTCTTTTGGAAACAATGCTACTCTGGGTAACA 2329
QY 2340 AATTGAGTCTCTGAGTGTACACGAAAAAATCAATGTGAGCCCAACAACCTTACAGATTAC 2399
Db 2330 AATTGAGTCTCTGAGTGTACACGAAAAAATCAATGTGAGCCCAACAACCTTACAGATTAC 2389
QY 2400 TCAAGACTGTCTGCTGGTTGCTGTTTACCCCTCGATCAAGGAGGAACTTCCACCAACT 2459
Db 2390 TCAAGACTGTCTGCTGGTTGCTGTTTACCCCTCGATCAAGGAGGAACTTCCACCAACT 2449
QY 2460 TCCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGCTTACCAACAAGGAACAACCT 2519
Db 2450 TCCCTAGTGAATCCCGGAATCTGCCATTTCTGGATGCTTACCAACAAGGAACAACCT 2509
QY 2520 CAATAACCATGCTTCCAAAAGAAAGATGTGGAAGATGGCAATGCCAATTTTGTAGGCAAG 2579
Db 2510 CAATAACCATGCTTCCAAAAGAAAGATGTGGAAGATGGCAATTTTGTAGGCAAG 2567
QY 2580 CATCAGGAATAGATTAGCAACCTATATACCTATTATTAAGTCTTTGGGTTTGACTCATT 2639
Db 2568 CATCAGGAATAGATTAGCAACCTATATACCTATTATTAAGTCTTTGGGTTTGACTCATT 2627
QY 2640 GAAGATGTAGAGCTCTGAAGCAGCTTGTAGGATTTAGATGCTACACAGCAGTATT 2699
Db 2628 GAAGATGTAGAGCTCTGAAGCAGCTTGTAGGATTTAGATGCTACACAGCAGTATT 2687
QY 2700 AAAATTTCAGGAGCAAGTCAACATCTTCTCTCTTTTAACTATGACAAAATGACTACT 2759
Db 2688 AAAATTTCAGGAGCAAGTCAACATCTTCTCTCTTTTAACTATGACAAAATGACTACT 2747
QY 2760 TCAGTACTTCACTCTTCAAGAAAAAATAAAACCTTAAAAAGTACTTTTGTGGGAGT 2819
Db 2748 TCAGTACTTCACTCTTCAAGAAAAAATAAAACCTTAAAAAGTACTTTTGTGGGAGT 2807
QY 2820 ATTTCTATTATATATACACGACCTTACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2879
```

```
Db 2808 ATTCTTATTATATATACACGACCTTACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2867
QY 2880 ATCAGATAATTCTAGTCAATTTGTACAAACTGATGAGTCACTGCAATCTCATATCTCTGG 2939
Db 2868 ATCAGATAATTCTAGTCAATTTGTACAAACTGATGAGTCACTGCAATCTCATATCTCTGG 2927
QY 2940 TGGATGCGATGCTTATTAAAGTGTGTTTGTGATAG-TGTCTCTCAAAAAAATAAAAAA 2998
Db 2928 TGGATGCGATGCTTATTAAAGTGTGTTTGTGATAGTGTCTCAAAAAAATAAAAAA 2987
QY 2999 AAAAAAATAAAAAA 3015
Db 2988 AAAAAAATAAAAAA 3004

RESULT 4
BC028384
LOCUS Homo sapiens, Similar to guanylate cyclase 1, soluble, alpha 3, 2657 bp mRNA linear PRI 25-APR-2002
DEFINITION clone MGC:33150 IMAGE:4792661, mRNA, complete cds.
ACCESSION BC028384
VERSION BC028384.1 GI:20306358
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2657)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 46 Row: a Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4504212.
FEATURES
Location/Qualifiers
1..2657
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:33150 IMAGE:4792661"
/tissue_type="Brain, hypothalamus"
/clone_lib="NIH MGC_96"
/lab_host="DH10B"
/note="Vector: pBluescript"
388..2460
/codon_start=1
/product="Similar to guanylate cyclase 1, soluble, alpha
3"
/protein_id="AAH28384.1"
/db_xref="GI:20306359"
/translation="MFTKLDKLTGECFPFSLAPGVNPESSEAAAGSSSCSKATM
PICQDIPKNTQESLPQKTSRSRVLLHTLAESICKLIPFEPELNVALQRTLAKHKI
KESKSLREDPEKTIAGQAVAGVPVEIKESLGEVFKICYEEDENILGVVGTGLK
DELNSFLLKQSSHCQAGKRGLEDASILCLDEDELHLYVFFPKRTTSLILPGI
TKAAHLVETEVEVSLPPCPHNDCEVFNQPLYLISVHMKSFTSPSPKQSSLV
IPTSLFCKTFPPHFDKMTILQFNGIRLRNRDRDQGNFNFEEFELPKINQT
```

ESGIMTMNMQFVVRVRWDNSVKKSRVMDLKGOMIYIVESSAILFLGSPCVRDLRD
FTGRLGLSDIPHNARDVVLGEQARAQDGLKRLGKATLEQAQHAQALEBEKKKT
VLLCSFFPCEAQOQWQVQAKFNSVMTLFSDFVGTATCSQSPLOVITMLNA
LYTRFOQCELDVYKVTIGDAYCVAGGLHKSDDTHAQIALMVKMMLSEVNSP
HGPP IKRIGLHSGSVFAGVVGKMPRYCLFGNNVTLANKFESCVPKR INVSPYYR
LLKDCPGEVFTPSREELPNFSEIPGICHLFDAYQOQINSKPCQKKRQVEDGNANF
LGRASGID*

BASE COUNT 742 a 619 c 669 g 627 t

ORIGIN

Query Match 78.98; Score 2378.2; DB 9; Length 2657;
Best Local Similarity 99.38; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 387 CTTGAGCTGCTGACAGTGACATGACATCCCACTTACCACTGCTTCAATGATAGTG 446
DB 251 CTTTACCCCGCGCTCGCGGTGGACATCCCACTTACCACTGCTTCAATGATAGTG 310
QY 447 GCTTCTGTTGTGCTGCTCATATAAGNACTACAGCTCATCAGGAGGATCGCAGAGG 506
DB 311 GCTTCTGTTGTGCTGCTCATATAAGAACTACAGCTCATCAGGAGGATCGCAGAGG 370
QY 507 TAAGAGACACCAACACCATCTTCTGACGAGCTCAAGGATCTCAAGATCACAGGAGGT 566
DB 371 TAAGAGACACCAACCATCTTCTGACGAGCTCAAGGATCTCAAGATCACAGGAGGT 430
QY 567 GTCTTCTTCTTCTTCTGACAGGCTCAAGTCTTCAAGGATCTTCAAGAGGAGGAGG 626
DB 431 GTCTTCTTCTTCTTCTGACAGGCTCAAGTCTTCAAGGATCTTCAAGAGGAGGAGG 490
QY 627 GAAGCTCAGAGCTCAAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 686
DB 491 GAAGCTCAGAGCTCAAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
QY 687 TACAAGAAAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 746
DB 551 TACAAGAAAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 610
QY 747 CAGAGAGTATTCGAACTGATTTTCCAGAGTTTGAAGGCTGAAATGTTGCTGCTGCT 806
DB 611 CAGAGAGTATTCGAACTGATTTTCCAGAGTTTGAAGGCTGAAATGTTGCTGCTGCT 670
QY 807 GAACATTTGCAAGCAACAAATAAAGAGAGGAGGAACTTTTGAAGAGAGAGAGCTTTC 866
DB 671 GAACATTTGCAAGCAACAAATAAAGAGAGGAGGAACTTTTGAAGAGAGAGAGCTTTC 730
QY 867 AAAAAACAATTCAGAGCAAGCAGTTCAGCAGGAGTTCAGTGGAGGTTTATCAAGAAT 926
DB 731 AAAAAACAATTCAGAGCAAGCAGTTCAGCAGGAGTTCAGTGGAGGTTTATCAAGAAT 790
QY 927 CTCTTGTGAGAGGTTTTTAAATATGTTACAGGAAGATGAAGACATCTTGGGGTGG 986
DB 791 CTCTTGTGAGAGGTTTTTAAATATGTTACAGGAAGATGAAGACATCTTGGGGTGG 850
QY 987 TTGAGGACCCCTTAAAGATTTTTTAAACAGCTTCAGTACCTTCTGAAACAGAGAGCC 1046
DB 851 TTGAGGACCCCTTAAAGATTTTTTAAACAGCTTCAGTACCTTCTGAAACAGAGAGCC 910
QY 1047 ATTGCCAAGAGAGGAGGAGGCTTGAAGACGCTTCAATCTATGCTGCTGATA 1106
DB 911 ATTGCCAAGAGAGGAGGAGGCTTGAAGACGCTTCAATCTATGCTGCTGATA 970
QY 1107 AGAGAGTATTTTCTACATGTTTACTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1166
DB 971 AGAGAGTATTTTCTACATGTTTACTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1030
QY 1167 TTCCGCGCATTAAGGAGGCTGCTCAGCTATATATGAAGAGGAGTGAAGTCTCT 1226
DB 1031 TTCCGCGCATTAAGGAGGCTGCTCAGCTATATATGAAGAGGAGTGAAGTCTCT 1090
QY 1227 TAATGCTCCTCCTTCCCAATGATTCAGGAGGTTTGTGAATCAGCCCTACTTGTGT 1286
DB 1091 TAATGCTCCTCCTTCCCAATGATTCAGGAGGTTTGTGAATCAGCCCTACTTGTGT 1150

QY 1287 ACTCCGTTTACATGAAGAAGCAAGCCATCCCTGTCCTCCCAAGCAACCCAGTCTCTCG 1346
DB 1151 ACTCCGTTTACATGAAGAAGCAAGCCATCCCTGTCCTCCCAAGCAACCCAGTCTCTCG 1210
QY 1347 TGTGATTTCCCATCGCTATTTCTGCAAGACATTTCCATTTCCATTTCTGCAAGAA 1406
DB 1211 TGTGATTTCCCATCGCTATTTCTGCAAGACATTTCCATTTCCATTTCTGCAAGAA 1270
QY 1407 ATATGCAATTTCTGCAATTTTGGCAATGGCATCAGAGGCTGATCAACAGAGAGACTTTC 1466
DB 1271 ATATGCAATTTCTGCAATTTTGGCAATGGCATCAGAGGCTGATCAACAGAGAGACTTTC 1330
QY 1467 AAGGAAAGCCTAATTTTGAAGAACTTTTGAATTTCTGACTTCCAAAATCAACAGACGT 1526
DB 1331 AAGGAAAGCCTAATTTTGAAGAACTTTTGAATTTCTGACTTCCAAAATCAACAGACGT 1390
QY 1527 TTAGCGGGATCATGACTATTTGTAATATGCAGTTTGTGTACAGTGAAGAGATGGACA 1586
DB 1391 TTAGCGGGATCATGACTATTTGTAATATGCAGTTTGTGTACAGTGAAGAGATGGACA 1450
QY 1587 ACTCTGTGAAGAACTCTTCAAGGTTTATGAGCTCAAGGCTCAAGGCTCAAGTCTACATTTG 1646
DB 1451 ACTCTGTGAAGAACTCTTCAAGGTTTATGAGCTCAAGGCTCAAGGCTCAAGTCTACATTTG 1510
QY 1647 AATCCAGTGAATCTTTGTTTGGGCTCACTTGTGTGAGAGATTAGAGATTTTACAG 1706
DB 1511 AATCCAGTGAATCTTTGTTTGGGCTCACTTGTGTGAGAGATTAGAGATTTTACAG 1570
QY 1707 GAGAGGGCTTACCTCTCAGACATCCCAATTCACATTCAGTGAAGGATGTGGTCTTAA 1766
DB 1571 GAGAGGGCTTACCTCTCAGACATCCCAATTCACATTCAGTGAAGGATGTGGTCTTAA 1630
QY 1767 TAGGGGAAAGCCCGAGCTCAAGATGGCTGAAGAGAGGCTGGGAGAGCTCAAGGCTA 1826
DB 1631 TAGGGGAAAGCCCGAGCTCAAGATGGCTGAAGAGAGGCTGGGAGAGCTCAAGGCTA 1690
QY 1827 CCCTTGAGCAAGCCCAAGCCCTTGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1886
DB 1691 CCCTTGAGCAAGCCCAAGCCCTTGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1750
QY 1887 GCTCCATATTTCCCTGAGGTTGCTCAGCAGCTGGCAAGGAGGAGGAGGAGGAGGAGGAG 1946
DB 1751 GCTCCATATTTCCCTGAGGTTGCTCAGCAGCTGGCAAGGAGGAGGAGGAGGAGGAGGAG 1810
QY 1947 AGAAGTTCAGTATGCTACCATGCTTCTTCAGACATCTTGGGTTTCACTGCTGCTGCT 2006
DB 1811 AGAAGTTCAGTATGCTACCATGCTTCTTCAGACATCTTGGGTTTCACTGCTGCTGCT 1870
QY 2007 CCAGTGTCCAGGCTGAGGCTCATCCATGCTCAATGCTCAATGCTCACTGCTGCTGCTGCT 2066
DB 1871 CCAGTGTCCAGGCTGAGGCTCATCCATGCTCAATGCTCAATGCTCACTGCTGCTGCTGCT 1930
QY 2067 AGCAGTGTGAGAGCTGGATGCTCAAGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2126
DB 1931 AGCAGTGTGAGAGCTGGATGCTCAAGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1990
QY 2127 CTGGGGGATTTACAAAGAGAGTACTACTGCTGCTTCAGATAGGCTGATGGCCCTGA 2186
DB 1991 CTGGGGGATTTACAAAGAGAGTACTACTGCTGCTTCAGATAGGCTGATGGCCCTGA 2050
QY 2187 AGATGATGAGCTCTCTGATGAGGTTATGTCCTCCCATGGAGAACCTATCAAGATGCGAA 2246
DB 2051 AGATGATGAGCTCTCTGATGAGGTTATGTCCTCCCATGGAGAACCTATCAAGATGCGAA 2110
QY 2247 TTGAGCTGCTGATCAGTTTGTGCTGGGCTGCTTGGAGTTAAATGCCCGCTTACT 2306
DB 2111 TTGAGCTGCTGATCAGTTTGTGCTGGGCTGCTTGGAGTTAAATGCCCGCTTACT 2170
QY 2307 GTCTTTTGGAAACAATGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2366
DB 2171 GTCTTTTGGAAACAATGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2230

```

QY 2367 AAATCAATGTCAGCCCAACAACTTACAGATTACTCAAGAGACTGTCTGGTTTCGGTTGTTA 2426
Db 2231 AAATCAATGTCAGCCCAACAACTTACAGATTACTCAAGAGACTGTCTGGTTTCGGTTGTTA 2290
QY 2427 CCCCTCGATCAAGGAGGAACTTCCACCAAACTTCCCTAGTGAATCCCGGAATCTGCC 2486
Db 2291 CCCCTCGATCAAGGAGGAACTTCCACCAAACTTCCCTAGTGAATCCCGGAATCTGCC 2350
QY 2487 ATTTTCTGATGCTTTACCAACAGGAACAACTCAAAACCAATGCTTCCAAAAGAAAGATG 2546
Db 2351 ATTTTCTGATGCTTTACCAACAGGAACAACTCAAAACCAATGCTTCCAAAAGAAAGATG 2410
QY 2547 TGGAGATGGCAATCCCAATTTTGGCAAGCAATCAGGAATAGATAGCAACCTATAT 2606
Db 2411 TGGAGATGGCAATCCCAATTTTGGCAAGCAATCAGGAATAGATAGCAACCTATAT 2470
QY 2607 ACCTATTTATAAGTCTTTGGGGTTTGAAGTCTTCAATTTGAAGATGTAGAGCTCTGAAAGCAC 2666
Db 2471 ACCTATTTATAAGTCTTTGGGGTTTGAAGTCTTCAATTTGAAGATGTAGAGCTCTGAAAGCAC 2530
QY 2667 TTTAGGATGTPAGATGGCTTAACAGCAGTATTAATAATTTTCAGGAGCCCAAGTCAACAATCT 2726
Db 2531 TTTAGGATGTPAGATGGCTTAACAGCAGTATTAATAATTTTCAGGAGCCCAAGTCAACAATCT 2590
QY 2727 TTCTCCTGTTTAAACATGACAAATGACTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 2786
Db 2591 TTCTCCTGTTTAAACATGACAAATGACTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 2650
QY 2787 AAAAAA 2793
Db 2651 AAAAAA 2657

RESULT 5
AK096588
LOCUS
DEFINITION
Homo sapiens cDNA FLJ39269 fis, clone OCBBF2010420, highly similar
to GUANYLATE CYCLASE SOLUBLE, ALPHA-1 CHAIN (EC 4.6.1.2).
ACCESSION
AK096588
VERSION
1.1
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens fetal brain cDNA to mRNA, clone_lib:OCBBF2
clone:OCBBF2010420.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,
Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
Otsuki, T., Sato, H., Ota, I., Wakamatsu, A., Ishii, S., Yamamoto, J.,
Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K.,
Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H.,
Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A.,
Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K.,
and Isogai, T.
NEDO human cDNA sequencing project
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 2627)
AUTHORS
Isogai, T. and Yamamoto, J.
TITLE
Direct Submission
JOURNAL
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1. .2627

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OCBBF2010420"
/tissue_type="brain"
/clone_lib="OCBBF2"
/dev_stage="fetal"
/note="cloning vector: pME18SFL3"
BASE COUNT 759 a 585 c 640 g 643 t
ORIGIN

```

```

Query Match 76.7%; Score 2312.6; DB 9; Length 2627;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2315; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 459 CAGTCTCATATAAGAACTACAGCTCATCAGGAGGAGATCGCAGCAGGGTAAGAGACACCA 518
Db 309 CAATCTCATATAAGAACTACAGCTCATCAGGAGGAGATCAGAGGGTAAGAGACACCA 368
QY 519 ACACCATGTTTGCAGCAAGCTCAAGGATCTCAAGATCAAGATCAAGAGAGTGTCTTCTCT 578
Db 369 ACACCATGTTTGCAGCAAGCTCAAGGATCTCAAGATCAAGATCAAGAGAGTGTCTTCTCT 428
QY 579 TACTGGCACCAGTCAAGTTCCTTAACGAGTCTTCAAGAGGAGGAGCAGCAGGAAGTCA 638
Db 429 TACTGGCACCAGTCAAGTTCCTTAACGAGTCTTCAAGAGGAGGAGCAGCAGGAAGTCA 488
QY 639 GCTGCAAGCAACCGTGGCCATCTGTCAAGATCTTCTGAGAGCAATCTTCAAGAGCAAT 698
Db 489 GCTGCAAGCAACCGTGGCCATCTGTCAAGATCTTCTGAGAGCAATCTTCAAGAGCAAT 548
QY 699 TTCCTCAAGAAAAACCGTGGCCATCTGTCAAGATCTTCTGAGAGCAATCTTCAAGAGCA 758
Db 549 TTCCTCAAGAAAAACCGTGGCCATCTGTCAAGATCTTCTGAGAGCAATCTTCAAGAGCA 608
QY 759 GCAAACTGATTTTCCAGAGTTTGAACGGCTTGAACGGCTTGAACGGCTTGAACGGCTTGA 818
Db 609 GCAAACTGATTTTCCAGAGTTTGAACGGCTTGAACGGCTTGAACGGCTTGAACGGCTTGA 668
QY 819 AGCAGCAAAATAAAGAAAGCAGGAAATCTTTGGAAGAGAGAGACTTTGAAAAACAATTTG 878
Db 669 AGCAGCAAAATAAAGAAAGCAGGAAATCTTTGGAAGAGAGAGACTTTGAAAAACAATTTG 728
QY 879 CAGAGCAAGCAGTTGCGCAGCAGGAGTTCAGTGGAGGTTTCAAGAGATCTTCTGTTGAG 938
Db 729 CAGAGCAAGCAGTTGCGCAGCAGGAGTTCAGTGGAGGTTTCAAGAGATCTTCTGTTGAG 788
QY 939 AGTTTTAAAAATATGTTACGAGGAGATGAAAAACATCTTGGGGTGGTTGGAGGCACCC 998
Db 789 AGTTTTAAAAATATGTTACGAGGAGATGAAAAACATCTTGGGGTGGTTGGAGGCACCC 848
QY 999 TTAAGAGATTTTAAACAGCTTTCAGTACCTTCTGAAACAGAGAGAGCCATTTGCCAAGA 1058
Db 849 TTAAGAGATTTTAAACAGCTTTCAGTACCTTCTGAAACAGAGAGAGCCATTTGCCAAGA 908
QY 1059 CAGGAAAAAGGGGCGAGGCTTGAGGAGCGCTTCAATCTATGCTTGGATAGGAGGATGATT 1118
Db 909 CAGGAAAAAGGGGCGAGGCTTGAGGAGCGCTTCAATCTATGCTTGGATAGGAGGATGATT 968
QY 1119 TTCTACATGTTTACTTCTTCTTCCCTTAAGAGAAACACCTTCCCTGATTTCTCCGGCATCA 1178
Db 969 TTCTACATGTTTACTTCTTCTTCCCTTAAGAGAAACACCTTCCCTGATTTCTCCGGCATCA 1028
QY 1179 TAAAGGAGCTGCTCACGTTATTTATATGAAACGAGAGTGGAGAGTGTCTGTTAAATGCCCT 1238
Db 1029 TAAAGGAGCTGCTCACGTTATTTATATGAAACGAGAGTGGAGAGTGTCTGTTAAATGCCCT 1088
QY 1239 GCTTCCATATGATGTCAGCAGGATTTGTGAATCAGCCCTTACTTGTGTTACTCGTTTACA 1298
Db 1089 GCTTCCATATGATGTCAGCAGGATTTGTGAATCAGCCCTTACTTGTGTTACTCGTTTACA 1148
QY 1299 TGAAGAGCACCAGCCATCCCTGTCCCCCAGCAAAACCCAGTCTCGCTGGTGGATTTCCA 1358
Db 1149 TGAAGAGCACCAGCCATCCCTGTCCCCCAGCAAAACCCAGTCTCGCTGGTGGATTTCCA 1208

```


HGEPIKMRIGLHSGSVFAGVGVKMPRYCLFCNNVTLANKFESCSVPKINVSPTTYR LKQDCPGVFTPSREELPPNFPSEIPGICHELFDAYQOQTNKSPCFQRKDDVGNANF LGKASGID*									
BASE COUNT	667	a	500	c	520	g	571	t	
ORIGIN									
Query Match	74.9%; Score 2258; DB 9; Length 2258;								
Best Local Similarity	100.0%; Pred. No. 0;								
Matches 2258; Conservative	0; Mismatches 0; Indels 0; Gaps 0;								
QY	507	TAAGAGACACCAACACATGTTCTGCACGAAGCTCAAGGATCTCAAGATCACAGGAGAGT	566						
DB	1	TAAGAGACACCAACACATGTTCTGCACGAAGCTCAAGGATCTCAAGATCACAGGAGAGT	60						
QY	567	GTCCCTTCTCCTTACTGCGACACAGGTCAAGTTCCTAAACGAGTTCACAGGAGGACGAG	626						
DB	61	GTCCCTTCTCCTTACTGCGACACAGGTCAAGTTCCTAAACGAGTTCACAGGAGGACGAG	120						
QY	627	GAAGCTCAGAGAGCTGCAAGCAACCGTCCCATCTGTCAAGACATTCCTGAGAAGAACA	686						
DB	121	GAAGCTCAGAGAGCTGCAAGCAACCGTCCCATCTGTCAAGACATTCCTGAGAAGAACA	180						
QY	687	TACAAGAAAGTCTTCCTCAAGAAACACAGTCGGAGCGAGTCTATCTTCACACTTTGG	746						
DB	181	TACAAGAAAGTCTTCCTCAAGAAACACAGTCGGAGCGAGTCTATCTTCACACTTTGG	240						
QY	747	CAGAGAGTATTGCAAACTGATTTTCCACAGAGTTGAACGGTGAATGTTGCACTTCAGA	806						
DB	241	CAGAGAGTATTGCAAACTGATTTTCCACAGAGTTGAACGGTGAATGTTGCACTTCAGA	300						
QY	807	GAACATTGCAAGCAACAAATAAAGAAAGCAGGAATCTTTGGAAAGAGAGACTTTG	866						
DB	301	GAACATTGCAAGCAACAAATAAAGAAAGCAGGAATCTTTGGAAAGAGAGACTTTG	360						
QY	867	AAAAACAATTCAGACCAAGCAGTTGCAGCAGGAGTTCAGTGGAGGTTATCAAGAAT	926						
DB	361	AAAAACAATTCAGACCAAGCAGTTGCAGCAGGAGTTCAGTGGAGGTTATCAAGAAT	420						
QY	927	CTCTTGGTGAAGAGGTTTTTAAATATGTTACGAGGAAGATGAAACATCTCTTGGGGTG	986						
DB	421	CTCTTGGTGAAGAGGTTTTTAAATATGTTACGAGGAAGATGAAACATCTCTTGGGGTG	480						
QY	987	TTGGAGGACCCCTTAAAGATTTTTTAAACAGCTTCAGTACCCCTTCGAAACAGAGCAGCC	1046						
DB	481	TTGGAGGACCCCTTAAAGATTTTTTAAACAGCTTCAGTACCCCTTCGAAACAGAGCAGCC	540						
QY	1047	ATTGCCAAGAAGCAGGAAAAAGGGCAGGCTTGAGGACGCCCTCATCTATGCTCGATA	1106						
DB	541	ATTGCCAAGAAGCAGGAAAAAGGGCAGGCTTGAGGACGCCCTCATCTATGCTCGATA	600						
QY	1107	AGGAGGATGATTTCTACATGTTTACTACTTCTTCCCTAAGAGACCCCTCCCTGATTC	1166						
DB	601	AGGAGGATGATTTCTACATGTTTACTACTTCTTCCCTAAGAGACCCCTCCCTGATTC	660						
QY	1167	TTCCCGGCGATCAAAAGGCGAGTGCPCAGCTATTATATCAAGCGGAAGTGGAGTCTCGT	1226						
DB	661	TTCCCGGCGATCAAAAGGCGAGTGCPCAGCTATTATATCAAGCGGAAGTGGAGTCTCGT	720						
QY	1227	TAATGCGCTCCCTGCTTCCATAATGATTCGAGCGAGTTTGTGAATCAGCCCTACTTGTGT	1286						
DB	721	TAATGCGCTCCCTGCTTCCATAATGATTCGAGCGAGTTTGTGAATCAGCCCTACTTGTGT	780						
QY	1287	ACTCCGTTACATGAAAGACACAGCCATCCCTGTCCCGCAGCAACCCCGAGTCTCCG	1346						
DB	781	ACTCCGTTACATGAAAGACACAGCCATCCCTGTCCCGCAGCAACCCCGAGTCTCCG	840						
QY	1347	TGCTGATTCACATCGCTATCTGCAAGACATTTCCATTCATTTTCATGTTTGACAAAG	1406						
DB	841	TGCTGATTCACATCGCTATCTGCAAGACATTTCCATTCATTTTCATGTTTGACAAAG	900						
QY	1407	ATATGACAATTTGCAATTTGGCAATGGCATCAGAAGGCTGATGAACAGGAGACTTTC	1466						
DB	901	ATATGACAATTTGCAATTTGGCAATGGCATCAGAAGGCTGATGAACAGGAGACTTTC	960						

QY	1467	AAGGAAAGCCTAATTTTGAAGATAACTTTTGAATTTCTGACTCCAAAAATCAACCAGACGT	1526						
DB	961	AAGGAAAGCCTAATTTTGAAGATAACTTTTGAATTTCTGACTCCAAAAATCAACCAGACGT	1020						
QY	1527	TTAGCGGGATCATGACTATGTTGAATATGCGAGTTTGTGTACGAGTGAGAGATGGGACA	1586						
DB	1021	TTAGCGGGATCATGACTATGTTGAATATGCGAGTTTGTGTACGAGTGAGAGATGGGACA	1080						
QY	1587	ACTCTGTGAAGAAATCTTCAAGGGTTATGACCTCAAGGCCAAATGATCTACATTTGTG	1646						
DB	1081	ACTCTGTGAAGAAATCTTCAAGGGTTATGACCTCAAGGCCAAATGATCTACATTTGTG	1140						
QY	1647	AATCCAGTGCATCTTGTGTTTGGGGTCAACCTGTGTGGACAGATTAGAAGATTTTACAG	1706						
DB	1141	AATCCAGTGCATCTTGTGTTTGGGGTCAACCTGTGTGGACAGATTAGAAGATTTTACAG	1200						
QY	1707	GACGAGGCTCTACCTCTCAGACATCCCAATTCACAATCACAATGCGACTGAGGGATGTGCTTAA	1766						
DB	1201	GACGAGGCTCTACCTCTCAGACATCCCAATTCACAATCACAATGCGACTGAGGGATGTGCTTAA	1260						
QY	1767	TAGGGAACAAGCCGAGCTCAAGATGGCTGAAGAGAGGCTGGGGAAGCTCAAGGCTA	1826						
DB	1261	TAGGGAACAAGCCGAGCTCAAGATGGCTGAAGAGAGGCTGGGGAAGCTCAAGGCTA	1320						
QY	1827	CCCTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAAAGAAAGACAGTAGACCTTCTGT	1886						
DB	1321	CCCTTGAGCAAGCCCAAGCCCTGGAGGAGGAGAAAGAAAGACAGTAGACCTTCTGT	1380						
QY	1887	GCTCCATATTTCCCTGTGAGTTGCTCAGCAGCTGTGGCAAGGGCAAGTTGCAAGCCA	1946						
DB	1381	GCTCCATATTTCCCTGTGAGTTGCTCAGCAGCTGTGGCAAGGGCAAGTTGCAAGCCA	1440						
QY	1947	AGAAGTTTCAATGTCACCATCTCTTCTCAGACATCGTTGGTTCACCTGCCATCTGCT	2006						
DB	1441	AGAAGTTTCAATGTCACCATCTCTTCTCAGACATCGTTGGTTCACCTGCCATCTGCT	1500						
QY	2007	CCAGTGTCAACGGCTGCAGGTCATCACCATGCTCAATGCATGTACACTCGCTTCGACC	2066						
DB	1501	CCAGTGTCAACGGCTGCAGGTCATCACCATGCTCAATGCATGTACACTCGCTTCGACC	1560						
QY	2067	AGCAGTGTGGAGAGCTGGATGCTCAAGGTGGAGACCATTTGGCGATGCTATTTGTAG	2126						
DB	1561	AGCAGTGTGGAGAGCTGGATGCTCAAGGTGGAGACCATTTGGCGATGCTATTTGTAG	1620						
QY	2127	CTGGGGATTAACAAAGAGAGTGTACTCATCTGTTTCAGATAGCGCTGATGCGCCCTGA	2186						
DB	1621	CTGGGGATTAACAAAGAGAGTGTACTCATCTGTTTCAGATAGCGCTGATGCGCCCTGA	1680						
QY	2187	AGATGATGAGCTCTCTGATGAAGTTATGTCTCCCATGGAGAACCTATCAAGATCGGAA	2246						
DB	1681	AGATGATGAGCTCTCTGATGAAGTTATGTCTCCCATGGAGAACCTATCAAGATCGGAA	1740						
QY	2247	TTGGACTGCACTCTGATCAGTTTTCGCGCTGCTGGAGTTAAATGCCCCCTTACT	2306						
DB	1741	TTGGACTGCACTCTGATCAGTTTTCGCGCTGCTGGAGTTAAATGCCCCCTTACT	1800						
QY	2307	GTCTTTTGGAAACAATGTCACTCTGGCTTAACAAATTTAGTCTCTGCAGTGTACACGAA	2366						
DB	1801	GTCTTTTGGAAACAATGTCACTCTGGCTTAACAAATTTAGTCTCTGCAGTGTACACGAA	1860						
QY	2367	AAATCAATGTACGCCCAACAATTACAGATTACTCAAGACTGCTCTGTTTTCGTTTAA	2426						
DB	1861	AAATCAATGTACGCCCAACAATTACAGATTACTCAAGACTGCTCTGTTTTCGTTTAA	1920						
QY	2427	CCCTCTGATCAAGGGAGAACTTCCACCAAACTTCCCTAGTGAATCCCGGATCTGCC	2486						
DB	1921	CCCTCTGATCAAGGGAGAACTTCCACCAAACTTCCCTAGTGAATCCCGGATCTGCC	1980						
QY	2487	ATTTTCTGGATGCTTACCAACAGGAACAAACTCAAAACCATGCTTCCAAAAGAAAGATG	2546						
DB	1981	ATTTTCTGGATGCTTACCAACAGGAACAAACTCAAAACCATGCTTCCAAAAGAAAGATG	2040						

QY 2547 TGAAGATGGCAATGCCAATTTTGGCAAGCATCAGGAATAGATTAGCAACCTATAT 2606
 DB 2041 TGAAGATGGCAATGCCAATTTTGGCAAGCATCAGGAATAGATTAGCAACCTATAT 2100
 QY 2607 ACCTATTATAAGTCTTTGGGTTGACTCATTTGAAGATGTTAGAGCCCTCTGAAGCAC 2666
 DB 2101 ACCTATTATAAGTCTTTGGGTTGACTCATTTGAAGATGTTAGAGCCCTCTGAAGCAC 2160
 QY 2667 TTTAGGATTTAGATGGCTTAACAAGCAGTATTAAATTTTTCAGGAGCCCAAGTCACAACTCT 2726
 DB 2161 TTTAGGATTTAGATGGCTTAACAAGCAGTATTAAATTTTTCAGGAGCCCAAGTCACAACTCT 2220
 QY 2727 TTTCTCTGTTTAAACATGACAAATGTTACTCATTCTAGT 2764
 DB 2221 TTTCTCTGTTTAAACATGACAAATGTTACTCATTCTAGT 2258

RESULT 7

BTGUCY73 3261 bp mRNA linear MAM 12-SEP-1993
 DEFINITION Bovine mRNA for large subunit of guanylate cyclase (EC 4.6.1.2).
 ACCESSION X54014
 VERSION X54014.1 GI:405
 KEYWORDS guanylate cyclase.
 SOURCE cow.

ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.

1 (bases 1 to 3261)
 Koesling,D. Issson

Direct Submission

Submitted (04-OCT-1990)

2 (bases 1 to 3261)

Koesling,D., Harteneck,C., Humbert,P., Bosserhoff,A., Frank,R.,

Schultz,G. and Bohme,E.

TITLE
 The primary structure of the larger subunit of soluble guanylyl
 cyclase from bovine lung. Homology between the two subunits of the
 enzyme

JOURNAL FEBS Lett. 266 (1-2), 128-132 (1990)

MEDLINE 90306336

PUBMED 1973124

FEATURES

source

Location/Qualifiers

1..3261

/organism="Bos taurus"

/db_xref="taxon:9913"

/tissue_type="adrenal medulla"

/clone_lib="lambda 2Ap"

291..2366

/note="guanylate cyclase large subunit"

/codon_start=1

/protein_id="CAA37960.1"

/db_xref="GI:406"

/db_xref="SWISS-PROT:P19687"

/translation="MFCALIKLQITGDCPFLLAPGVDPREPLGEATGSGPASTPGQ
 PGVCPGVPDKNPGRPLRRNTSRVLTHTLAESICKLIPFEFERLNLAQRTLAHK
 IKENSLREDEKIVDQAIAGVPVEIKESLGEELFKICYEDEYILGVGVTL
 KDFNSFTLLKSCSHCOAEKKREFEDASILCLDKDPDLYVYFFPKRITSLILPG
 IIAAARILYETEHEVSVSTSPRHOCDREFVQPCELYSVHRSRPHPPKPVSSL
 VPASLFCKTFPEFVMDLRDMSILQILGHGIRRLMSRDVQKPFDEYEPILPKISQ
 TFSGLTMLNQPLVRRVNDNSMKSSRVNDLKGOMIYVESSSILFLGSPVDRL
 DFTGRGLYSDIPFHNALRDVLIGQARQDGLKRLKATLQALQALQALQALQALQ
 TVLLCSIFPEVPAHMOGHVQAKRFNVTMLFSDIVGFTALCSQCSPLOVITMLN
 ALYTRDQCGLDVKVETIDGAYCVAGLKHESDTHAVQIALMALKMLKMLKMLKMLK
 PHGEIKMRILHSGSVFAGVGVKMPRYCLFGNNVTLANQKLFESCSVPKINVSPTTY
 RLKDCPGFVETPSREELPPNFPSPDIPGICHELAYQQGTTSKPWFQKQDVEANAN
 FLKASGID"

BASE COUNT 845 a 789 c 814 g 813 t

ORIGIN

Query Match

Best Local Similarity 53.8%; Score 1620.6; DB 4; Length 3261;

Matches 2006; Conservative 0; Mismatches 479; Indels 28; Gaps 7;

QY 511 AGACACCAACACCATGTTTTCACAGAGCTCAAGGATCTCAAGATCAGAGGAGTGTC 570
 DB 278 AGGAGCAATCACCATTGTTTTCGCGAAGCTGAAGACCTCAGATCAGGGGACTGCC 337
 QY 571 TTTTCTCTTACTTGGCAGCAGGTCAAGTTCCTAAGAGTCTTTCAGAGGAGGAGGAG 630
 DB 338 TTTTCTCTCTCTGCGCCCGGGGAGGTCCCGCAGAGGCGCTTGGGGGAAGCCACAG 397
 QY 631 CTCAGAGAGCTGCAAAAGCAACCGTGCC--CATCTGTCAAGACATCTCTCAGAGAC 687
 DB 398 TGGGCCAGCATCGACCCAGGCGAGCCAGCGGTCTGTCCAGGTGTCCCCACAGAACCC 457
 QY 688 ACAAGAAAGTCTTCTCTCAAAAGAAAACCACTCGGAGCGGAGTCTATCTTCCACACT 747
 DB 458 TCCCGGAAGCTTCCCGGAGGAAGACCAACCGCGGTCTATCTTCCACTTTGGC 517
 QY 748 AGAGATATTGTCAAACTGATTTTCCAGAGTTTGAACGGCTGAATGTTGCATCTCAGAG 807
 DB 518 AGAGATATTGTCAAACTGATTTTCCAGAGTTTGAACGGCTTGAACCTTCCCTTCCAG 577
 QY 808 AACATTGGCAAGCACAATAAAGAAAGCAGGAAATCTTTGGAAGAGAGAGACTTTGA 867
 DB 578 AACATTGGCAAGCACAATAAAGAAAGCAGGAAATCTTTGGAAGAGAGAGACTTTGA 637
 QY 868 AAAACAATTTGACAGCAAGCAGCTTGCAGCAGGAGTTTCCAGTGGAGGTTTATCAAGA 927
 DB 638 AAAAATAGTTTGTGACCAAGCAATTTGCAGCAGGAGTTCCAGTGAGATCATCAAGA 697
 QY 928 TCTTGTGGAAGAGGTTTAAATATGTTTACGAGGAGGATGAAACATCTCTGGGGTGGT 987
 DB 698 TCTTGTGGAAGAGGTTTAAATATGTTTAAATATGTTTAAATATGTTTAAATATGTT 757
 QY 988 TGGAGGACCCCTTAAAGATTTTAAACAGCTTCAGTACCTCTCTGAAACAGAGCAGCCA 1047
 DB 758 TGGAGGACCCCTTAAAGATTTTAAATAGCTTTCAGCAGCCCTCTGAAACAGAGCAGCCA 817
 QY 1048 TTCCCAAGAGCAGGAAAGAGGGCAGGCTTTCAGGAGGAGGCTTCCATTTATGCTGTGATA 1107
 DB 818 CTGCTCAAGAGCAGAAAGAGGCGAGGTTTGGAGGAGGATCCATCTCTATGCTGTGATA 877
 QY 1108 GGAGGATGATTTTCTACATGTTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1167
 DB 878 AGATCCCGATGCTTATATGTTTACTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 937
 QY 1168 TCCCGGATCATTAAGGAGCAGCTGCTCAGCTATTATATGAAACGAGGAGTGAAGTGTGCT 1227
 DB 938 CCCCAGGATCATTAAGGAGCAGCTGCTCAGCTATTATGAGACGAGGAGTGGAGGTTTCT 997
 QY 1228 AATGCTCTCTGCTTCCATTAATGATTGACGAGGTTTGTGAATCAGCCCTTACTTGTGTA 1287
 DB 998 GAGCGCTTCTGCTTCCACAGGAGTCCCGTGGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1057
 QY 1288 CTCGGTTCATCAAAAGCAGCAAGCATCCCTGTCCCGCAGCAAAACCCAGTCTCGCT 1347
 DB 1058 CTCGGTTCATCAAGAGGAGCCCGGCGCAGCCCGCCCTTGGGAAACCCCGTGTCTCGCT 1117
 QY 1348 GGTGATTCCACATCGCTATTCTGCAAGACATTTTCCATTTCCATTTTATGTTTGCACAA 1407
 DB 1118 GGTGATTCCCGCTCGCTCTTCTGCAAGACTTTCCTTTCCATTTTATGTTTGCACAGA 1177
 QY 1408 TATGACAAATTCGCAATTTGGCAATGGCATGAGAGGCTGATGAACAGGAGAGACTTTCA 1467
 DB 1178 TATGAGCATCTCAGCTCGGCCACGCATCAGCGGCTGATGAGCAGGAGAGAGCGTGCA 1237
 QY 1468 AGGAAGCCCTTAATTTTGAAGAATACCTTGAATTCGACTCTCAAAATCAACAGAGCTT 1527
 DB 1238 AGGAAGCCCTTAATTTTGAAGAATACCTTGAATTCGACTCTCAAAATCAACAGAGCTT 1297
 QY 1528 TAGCGGATCATGACTATGTTGAATATGCAAGTGTGTTGACGAGTGGAGGAGTGGGACAA 1587
 DB 1298 CACGGCATTAAGCCATGCTGAACATGCAAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1357

FLNSPFTLLKQSHCOBAERRGLEDAISILCDKDDQLNVYFFPKRTALLPGII
KAAARILYSHVSEVSLMPPCFRSECTEFVNPQYLLSVHVKSPKSLSPGKPOSSIVI
PTSLFGLFFHMLDLALQLGNGIRLVRNRFQKGNFEFFFIILTPRINOTF
SGTWMLNMOFVIRNRWDNLVKKSSRVMDLKGOMIYIVESSAILFLGSPCVDRLEDF
TGRGLYLDIPINLRDVLVIGEOAPADGLKRLKGLKATLEHAHQALEEKKETV
DLCSIFPSVAQOLWQGLVQAKKFNENVTMLSDIVGFTAICSCSPQLQVITMLNAL
YTFPDQCGELDYKVEYIGDATVCYAGGLHRESHTAVQJALMALMMELSNVSPH
GEPFKMRIGLHSGSVFAGVGVKMPRYCLFNVTLANKFESCVPRKINVSPTVRL
LKDCPGFVTPRREBELPNFPNDIPGICHFLDAYOHQGNPSKPFQOKDAEDGNANF
LGRASGVD*

BASE COUNT 1340 a 1069 c 1042 g 1324 t
ORIGIN

Query Match 52.2%; Score 1573.6; DB 10; Length 4775;
Best Local Similarity 76.8%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 589; Indels 25; Gaps 8;

QY	374	CTGCACCCCTGTCGCTGAGTGCCTGACAGTGACATGACATCCAGTTACAGTGTCTT	433
DB	314	CTGCACCGGATGCCCTCGGCTCCCGTGACCGCATCATGCTGGGCCACTCGTGTCTT	373
QY	434	TGAATTGATAGTGGCTCTCTTTGTGCTCAGTCTCATATAAGAACTACAGCTCATCAGGAGGA	493
DB	374	TGAGTCAGTAGAGCAGATCTTCATCAGTC---CACATCAACACCGGCTTAATAGGAGGA	430
QY	494	GATCGCAGCAGGTAAGAGACACCAACCATGTTCTGACGAGCTCAAGGATCTCAAG	553
DB	431	AACCACTGC---CAAGCTCCAGAACACCATGTTCTGCAGGAAGTTCAAGATCTCAAG	486
QY	554	ATCACAGGAGTGTCTTCTCTCTTACTTGGCCAGGTCAGTCAAGTCTCAACGAGTCTTCA	613
DB	487	ATCACGGGGAGTGTCTCTTCTCTTACTTGGCTCTGCTGAGTGTCTTACGGGACCAAYA	546
QY	614	GAGGAGCAGCAGGAGCTCAGAGAGCTGCAAGCAACCGTGCCCACTCTCTCAAGACATT	673
DB	547	GAGGAGTGCCCGAGTCTCTGAGAGCTGCCAGGCCACTCTGCCACCTGCCAGGAGTTT	606
QY	674	CTTGAGAAACATACAGAAAGTCTTCTCAAGAAACCAAGTGGAGCCGAGTCTAT	733
DB	607	GCTG---AGAATGAGAGGAGTCAACCCCAAGAAAGAACAGCGCAACAGAGTCTAC	663
QY	734	CTTCACACTTTGGCAGAGTATTTGCACACTGATTTTCCAGAGTTTGAACGGCTCAAT	793
DB	664	CTACACCTTGGCAGAGATATTTGCAAGCTCATTTTTCCAGAGTTTGGAGACTGAAC	723
QY	794	GTTGCACTTCAGAGAACATTGGCAAGCAAAAATAAAGAAAGCAGGAATCTTTGGAA	853
DB	724	CTTGCACTTCAGAGACATTGGCAAGCACAATAAAGCAACAGGAATCTTTCAGAA	783
QY	854	AGAGAGACTTTGAAAACAAATGGCAGACCAAGCAGTTGCAGCAGAGTTCCAGTGGAG	913
DB	784	AAAGAGACCTTGAAGAATAATTTGCAGAAAGAACAAATTCGCGCAGGTGTCCAGTGGAG	843
QY	914	GTTATCAAGAAATCTCTTGTGTGAAGAGTTTAAATATGTTACGAGGAAGATGAAAC	973
DB	844	GTGCTCAAGACACTCTGGGTGAGGAACTGTTCAAGATCTGTTATGAGGAAGATGAGCAC	903
QY	974	ATCCTTGGGTGGTGGAGCCCTTAAAGATTTTTTAAACAGCTTCAGTACCCCTCTG	1033
DB	904	ATCCTGGGCTGTGGCGCACCTTGAAGGACTTTCTCAACAGCTTCAGCAGCTCTCTG	963
QY	1034	AAACAGAGCAGCAATTCGCAAGAACGAGGAAAGGGGCGAGGCTTGGAGCGCTCCATTT	1093
DB	964	AGCAGAGCAGCAGCTGCGAGGAGGAGGCGGGGCGACTGGAAGATGCTCCATC	1023
QY	1094	CTATGCTGATGAAGGAGGATGATTTCTACATGTTTACTACTTCTTCCCTAAGAGAAC	1153
DB	1024	TTGTGCTGCAAGGACGAGACTTTCTAAATGTTTACTACTTCTTCCGAGAGAAC	1083
QY	1154	ACCTCCCTGATTTCTCCCGGCATCAATAAGGCGAGCTGCTCACGATATATATGAACGGAA	1213
DB	1084	ACAGCCCTGCTCTCCCGGCTATCAUTAAAGCGGCTGCTGCCATCTGTACGAAAGCCAC	1143

QY	1214	GTGGAAGTGTCTTAATGCTCCCTCCCTCCATTAATGATTTGACGAGGATTTGTGAATCAG	1273
DB	1144	GTGGAGTGTCTCCCTGCTCCCTCCCTCCGAAAGTGAAGTGAATTCGTGAACCAAG	1203
QY	1274	CCCTACTTGTGTACTCCGTTTACATGAAAGAACCAACCCCTGCTCCCTCCAGCAAA	1333
DB	1204	CCCTATTTGCTTACTCCGTTCCAGTGAAGAGCACCAGGCTTCTCTGTCTCCAGGCAAA	1263
QY	1334	CCCAGTCTCTGCTGTGATTTCACATGCTCTTCTGCAAGACATTTCCATTCCTCATTTTC	1393
DB	1264	CCCAGTCTCTGCTGTGATTTCCTACTTCTCTGCAAACTTTCCCATTCCTCATTTTC	1323
QY	1394	ATGTTTGACAAAGATATGACAAATTTCTCAATTTGGCAATGGCATGCAAGGCTGATGAAC	1453
DB	1324	ATGCTGGACCGTGACCTGGCCATTCTGCAGCTGGGTAATGGCATGCAAGGCTGGTGAAC	1383
QY	1454	AGGAGAGACTTTCAAGGAAGGCTTAATTTTGAAGAATACTTTGAAATTTGACTCCAAA	1513
DB	1384	AAGAGGAGCTTCCAGGGAAGCCCAATTTTGAAGAGTCTTTTGAATTTTAACTTCCAAA	1443
QY	1514	ATCAACGACAGCTTTAGCGGATCATGACTATGTTGAATATGAGTGTGTGTACAGTGT	1573
DB	1444	ATCAACGACACTTTAGCGCATCATGACAAATGTTGAATATGAGTGTGTGTACAGGTTG	1503
QY	1574	AGGAGTGGGACAACTCTGTGAAGAAATCTTCAAGGTTTATGACCTCAAGGCCAAATG	1633
DB	1504	AGGAGTGGGATTAATCTAGTGAAGAAATCTTCAAGGTTTATGACCTCAAGGCTCAATG	1563
QY	1634	ATCTACATTTGTAATCCAGTGAATCTTTTGGGGTCACTCTGTTGGGACAGATTA	1693
DB	1564	ATCTACATCTTGAATCCAGTGCCTCTTTCTTAGGGTGGCGGTGTGTGGACAGATG	1623
QY	1694	GAAGATTTTACAGGAGGAGGCTCTACTCTCAGACATCCCAATTCACATGCACTGAGG	1753
DB	1624	GAAGATTTACAGGAGGAGGCTCTACTCTCTGACATCCCGATTCATATGSCCTGAGG	1683
QY	1754	GATGTGCTTAAATAGGGAAACAGCCGAGCTCAAGATGGCTGGAAGAGAGGCTGGGG	1813
DB	1684	GATGTGCTTGAATAGGGAAACAGGCGGCTCAAGATGGCTCAAGAGAGGTTGGGA	1743
QY	1814	AACTGAAGCTACCTTGGAGAGCCCAACAGCCCTGGAGGAGGAGGAGGAGGAGGAGCA	1873
DB	1744	AACTGAAGCAACCTTGGAGAGTGGCCCAAGCCCTGGAGGAGGAGGAGGAGGAGGAGC	1803
QY	1874	GTAGACTCTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCTGTGGCAAGGSCAA	1933
DB	1804	GTAGACTCTGTGCTCTATCTTTCCCTGTGAGTTGCTCAGCAGCTCTGSCAGGACAA	1863
QY	1934	GTTGTGCAAGCCCAAGAGTTTCAATGTCACCATGCTCTTCTCAGACATCTGTTGGGTTT	1993
DB	1864	ATTGTCAAGCCCAAGAGTTCAATGAGGTCACTGCTCTCTCAGATATCTGTAGGTTT	1923
QY	1994	ACTGCCATCTGTCTCCAGTCTCACCGTCCAGGTCATACCATGCTCTCAATGCACTGTAC	2053
DB	1924	ACTGCCATCTGTCTCAGTGTCCCTCTGCGAGTCACTACCATGCTCAACGCTCTCTAT	1983
QY	2054	ACTCGTGTGACACGAGTGTGGAGAGTGGATGTCTTACAAGGTGGAGACCATTTGGCGAT	2113
DB	1984	ACTCGTGTGACCAACAGTGTGGAGAGTGGATGTCTTACAAGGTGGAGACCATCTGGAGT	2043
QY	2114	GCTATTTGCTAGCTGGGGATTTACAAAGAGAGTGAATCTCTCTGTCAGATAGCG	2173
DB	2044	GCATATTTGTGCGAGGGGATTTGCACAGAGAGTGAACACATGCTGTCCAGATAGCA	2103
QY	2174	CTGATGCCCTGAAGATGATGAGGCTCTCTGATGAAGTTATGTCTCCCCATGGAGACCT	2233
DB	2104	CTGATGCCCTGAAGATGATGAGGCTCTCAATGAGGTCTATGTCTCCCCACGAGACCT	2163
QY	2234	ATCAAGATGGAAATTTGAGTGCATCTGTGAATAGTTTTTGTCTGGCTGTGTGTGAGTTAAA	2293
DB	2164	ATCAAGATGGAAATTTGAGTGCATCTTGTGGTCTAGTGTCTGTGGAGTGTGTGGAGTGAAG	2223
QY	2294	ATCCCCCGTTACTTCTTTTTTGGAAACAATGTCATCTCTGCTCAACAAATTTGAGTCTTGC	2353

Db	2224	ATGCCCGGATATTCCTGTTGGTGAACAATGTCACCTGCGCTAACAAATTTGAGTCTGCTC	2283
QY	2354	AGTGACCAAGAAATCAATGTCAGCCCAACAACCTTACAGATTACTCAAGACTGCTCT	2413
Db	2284	AGTGTGCTCGGAATAATGTCAGCCCAACAACCTTACTCAAGACTGCTCT	2343
QY	2414	GGTTTCGTGTACCCCTCGATCAAGGAGGAACTTCCACCAAACTTCCCTAGTGAATC	2473
Db	2344	GGTTTGTGTTCAACCCGAGATCGAGGAGGAGCTTCCACCAAACTTCCCTAGTGAAT	2403
QY	2474	CCGGGAATCTGCCATTTCTGGATGCTTACCAA---CAAGGACAAACCTCAAAACCATGC	2530
Db	2404	CTGGGAATCTGTCACCTTCTGGATGCTTACCAATCAAGGACCTAATTTCAACCATGG	2463
QY	2531	TTCCAAAAGAGATGGAAGATGCAATGCCAATTTTTAGGCAAGCATCAGGAATA	2590
Db	2464	TTCCAGCAGAAAGATGCAAGATGGAACGCTAATCTTAGGCAAGCATCAGGGTA	2523
QY	2591	GATTAGCAACCTATATACCTATTATAGTCTTTGGGGTTTGACTCATTTGAAGATGTGA	2650
Db	2524	GATTAGTAGGCGCATGTTCTGTGTTGATGCTTTGAGGGGTGCTCAGAATCTCTGAA	2583
QY	2651	GAGCCTCTGAAGACCTTTAGGATGTTAGTGGCTAACAAAGCAGTATTAAATTTGAGS	2710
Db	2584	ATGACTTTAGGATGCAAGATATAAAACTGTGTTTAAATTCGAGGGGTGAACCCAGT	2643
QY	2711	AGCCAAAGTCACAATCTTCTCTGTTTAAACATGACAAATGTACTCCTTCACTTCACTTCA	2770
Db	2644	TTCTCTGTTTCCATTTAAGATGGAATAAAAGTATTCAATGCTTCAGCGCTTTCAA	2703
QY	2771	GC-TCTTCAAGAAAAAACCCTTAAAGTACTTTTG---TGGAGTATTTCFA	2826
Db	2704	CCTTCTCTCTATTAAAGAAACAAACCTCAAAAGTACTGTGACTTTTGGGACATAGC	2763
QY	2827	TTATATAACGACGACTTACTACTGTAACAAATTCAGACCTTGTACATATATCAGAT	2886
Db	2764	TTGTGTTAGAGCAGTACTACTGCTGCAATCCAAATCCACACAGTGTATATTCAGAC	2823
QY	2887	AATTGTAAGTCAATGTA-----CAAACTGATGAGTACCTGCAATCTCATATCCCTGGTG	2941
Db	2824	AATTGTAAGTCAATGTAAGTGTCAAAAGTAAATGTAATGTAATGTAATGTAATGTAATG	2883
QY	2942	GAATGCCATGTTTAAAGTGTGTTGTAAGTGTGCTGCAAAATTCAGACCTTCAAAATTC	3001
Db	2884	AACTG---TGGTTATTACAGTGTGCTGTTGGTGTGCTGTAATCTCTGAATAGAA	2940
QY	3002	AAAAA 3007	
Db	2941	ATGACA 2946	

RESULT 9
AF297082
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2445 bp mRNA linear ROD 01-OCT-2000
Mus musculus soluble guanylyl cyclase alpha 1 subunit mRNA,
complete cds.
AF297082
AF297082.1 GI:10442713
Mus musculus.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2445)
Sharina, I.G., Krumenacker, J.S., Martin, E. and Murad, F.
Genomic organization of alpha 1 and beta 1 subunits of the mammalian
soluble guanylyl cyclase genes
Proc. Natl. Acad. Sci. U.S.A. 97 (20), 10878-10883 (2000)
20461464
10984516
2 (bases 1 to 2445)
Krumenacker, J.S., Sharina, I.G., Martin, E.S. and Murad, F.

TITLE	Direct Submission
JOURNAL	University (17-AUG-2000) Integrative Biology and Pharmacology, University of Texas Houston Medical School, 6431 Fannin St., Houston, TX 77030, USA
FEATURES	Location/Qualifiers
source	1..2445
CDS	/organism="Mus musculus" /strain="BALB/c" /db_xref="taxon:10090" 286..2361 /codon_start=1 /product="soluble guanylyl cyclase alpha 1 subunit" /protein_id="AAG17446.1" /db_xref="GI:10442714"
BASE COUNT	635 a 624 c 639 g 547 t
ORIGIN	
Query Match	51.5%; Score 1552; DB 10; Length 2445;
Best Local Similarity	82.3%; Pred. No. 0;
Matches 1810; Conservative	0; Mismatches 380; Indels 10; Gaps 2;
QY	420 GTTACCAGTCTCTTCAATTTGATGCTGCTTCTTGTTCAGTCTCATATAAGAACTACA 479
Db	189 GTCACCTCTCTCTTCAATTTGATGCTGCTTCTTGTTCAGTCTCATATAAGAACTACA 248
QY	480 GCTCATCAGGAGGATCGCAGCAGGTGAAGACACCAACCATCTTCTGCACCAAGC 539
Db	249 AGTCAGAAAGAACCCACTGC-----CAAGCTCCAGGAACACCATCTTCTGCAGGAAGT 301
QY	540 TCAAGCATCTCAAGATCACAGGAGAGTGTCTTCTTCTCTTCTTCTTCTTCTTCTTCT 599
Db	302 TCAAGCATCTCAAGATCACAGGAGAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 361
QY	600 CTAACGAGTCTTTCAGAGGAGGAGCAGGAGGCTCAGAGAGCTCAGAGCAACCGTCGCCA 659
Db	362 CTAAAGGAGCAGCAGGAGGAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 421
QY	660 TCTGTCAAGACATCTCTGTGAGAGAACATACAGAAAGTCTTCTCTCAAGAAACCAAGTC 719
Db	422 TCTGTCCAGTACTTCTGTGAGAGAACATCCAGAGGAGTCTCCCCCAAGAAAGACAGCC 481
QY	720 GGAGCCGAGTCTATCTTCACACTTTGCGAGAGAGTATTTGCAAACTGATTTTCCCAGAGT 779
Db	482 GCAACGAGTCTACCTGCAACACCTGCGCAGAGAGTATTTGCAAGCTCATCTTCCCAGAGT 541
QY	780 TTGAACGCTGATGTTGCACTTTCAGAGACATTTGCAAGAGCAGCAAAATAAGAAAGACGA 839
Db	542 GTGAGGAGTGAACCTTGTCACTTTCAGAGAACCTTTGCAAGAGCATAAATAAGAAAGAACCA 601
QY	840 GGAAATCTTTGAAAGAGAGAGTCTTTGAAAAACAATTCGAGAGCAAGCAGTGTGCAGCAG 899
Db	602 GGAAATCTTCAGAAAAAGAGAGCTTTGAAAAAATATTCGAGAGAGAGCAATTTGCAGCAG 661
QY	900 GAGTTCAGTGGAGGTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAATATGTTTACG 959
Db	662 GTCCCCAGTGGAGGCGCTCAAGAGCTCTCTGGCGAGGAGCTGTTCAAGATCTGCTATG 721
QY	960 AGGAAGATGAAACATCTCTTGGGCTGGTGGAGGACCCCTTAAAGATTTTTTAAACAGCT 1019
Db	722 AGGAAGATGAGCAATTTTGGGCTGGTGGCGGACCCCTGAAGAGACTTTCTTAAATAGCT 781

Db 2293 CGGAAATCAATGTACGCCCCACACATACAGGTACTCAAGACTGTCTGGCTTTGTG 2352
QY 2423 TTTACCCCTCGATCAAGGGAGGAATCTCCACCAAACTTCCCTAGTGAATCCCGGAATC 2482
Db 2353 TTTACCCCGAGATCGAGGGAGGACTTCCACCAAACTTCCCTAGTGACATCTCTGGATC 2412
QY 2483 TGCCATTTTCTGGATGCTTACCAGAA--CAAGGAACAAACTCAAAACCATCTTCCAAAG 2539
Db 2413 TGTACATTTCTGGATGCTTATCAACATCAAGGACCTAATTTCTAAACCATGTTTCCAGCAG 2472
QY 2540 AAGATGCTGAAGATGCAATGCAATTTTATAGGCAAGCATCAGGAATAGATTAGCAA 2599
Db 2473 AAGATGCAAGATGGAACGCTAATCTTAGGCAAGCATCAGGGGTAGATTAGTGA 2532
QY 2600 CCTATATACCTATTAT 2616
Db 2533 GCCGCAATGTTCTGTGT 2549

RESULT 11
AF233746S1 1571 bp DNA linear PRI 16-APR-2000
LOCUS Homo sapiens soluble guanylate cyclase large subunit gene, exon 3.
DEFINITION AF233746
ACCESSION AF233746.1 GI:7576897
KEYWORDS 1 of 6
SEGMENT Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Zhou,Y., Zheng,J.B., Gu,X., Li,W. and Saunders,G.F.
AUTHORS A novel Pax-6 binding site in rodent Bl repetitive elements:
TITLE coevolution between developmental regulation and repeated elements?
JOURNAL Gene 245 (2), 319-328 (2000)
MEDLINE 20183694
PUBMED 10717483
REFERENCE 2 (bases 1 to 1571)
AUTHORS Zhou,Y.H.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-2000) Biochemistry and Molecular Biology, The
University of Texas M. D. Anderson Cancer Center, 1515 Holcombe
Bldg., Houston, TX 77030, USA
FEATURES Location/Qualifiers
source 1..1571
/organism="Homo sapiens"
/db_xref="taxon:9606"
exon 378..1087
BASE COUNT 517 a 250 c 310 g 494 t
ORIGIN

Query Match 23.88; Score 716.2; DB 9; Length 1571;
Best Local Similarity 99.68; Pred. No. 8.2e-184;
Matches 718; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 897 CAGGAGTCCAGTGGAGGTTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAATATGTT 956
Db 375 CAGGAGTCCAGTGGAGGTTATCAAGAAATCTCTTGGTGAAGAGGTTTTTAAATATGTT 434
QY 957 ACAGGAGATGAACATCTTGGGTGGTGGAGGACCCCTTAAGATTTTTTAAACA 1016
Db 435 ACAGGAGATGAACATCTTGGGTGGTGGAGGACCCCTTAAGATTTTTTAAACA 494
QY 1017 GCTTCAGTCCCTTCTGAACAGAGCAGCCATTCGCAAGAGCAGGAAGAGGCGAGGC 1076
Db 495 GCTTCAGTCCCTTCTGAACAGAGCAGCCATTCGCAAGAGCAGGAAGAGGCGAGGC 554
QY 1077 TTGAGGAGCGCTCCATCTATGCTGTGAAGAGGAGGATGATTTTCTACATGTTTACTACT 1136
Db 555 TTGAGGAGCGCTCCATCTATGCTGTGAAGAGGAGGATGATTTTCTACATGTTTACTACT 614

QY 1137 TCCTCCCTRAGAACACACCTCCTGATCTTCCCGGCATCATAAAGGACGCTGCTCACG 1196
Db 615 TCCTCCCTRAGAACACACCTCCTGATCTTCCCGGCATCATAAAGGACGCTGCTCACG 674
QY 1197 TATTATATGAACGGAAGTGAAGTGTGTTAATGCTTCCCTGCTTCCATTAATGATGCA 1256
Db 675 TATTATATGAACGGAAGTGAAGTGTGTTAATGCTTCCCTGCTTCCATTAATGATGCA 734
QY 1257 GCAGGTTGTGATGACCCCTACTTGTGTACCTCCGTTACATGAAAGCACCAGCCAT 1316
Db 735 GCAGGTTGTGATGACCCCTACTTGTGTACCTCCGTTACATGAAAGCACCAGCCAT 794
QY 1317 CCCTGTCCCCCAGCAACCCAGTCTCGCTGTGTGATTTCCACATCGCTATTCTGCAAGA 1376
Db 795 CCCTGTCCCCCAGCAACCCAGTCTCGCTGTGTGATTTCCACATCGCTATTCTGCAAGA 854
QY 1377 CATTTCCATTCATTTGTTGACAAAGATATGACAATTTCTGCAATTTGGCAATGGCA 1436
Db 855 CATTTCCATTCATTTGTTGACAAAGATATGACAATTTCTGCAATTTGGCAATGGCA 914
QY 1437 TCAGAAAGGCTGATGAACAGGAGAGCTTTCAAGGAAAGCCCTAATTTTGAAGAATATCTTG 1496
Db 915 TCAGAAAGGCTGATGAACAGGAGAGCTTTCAAGGAAAGCCCTAATTTTGAAGAATATCTTG 974
QY 1497 AAATTCTGACTCCAAAATCAACCAGACGTTTACGGGATCATGACTATCTTGAATATGC 1556
Db 975 AAATTCTGACTCCAAAATCAACCAGACGTTTACGGGATCATGACTATCTTGAATATGC 1034
QY 1557 AGTTTGTGTACAGTGAAGAGATGGACAACTCTGTGAAGAATCTTCAAGGGTTATGG 1616
Db 1035 AGTTTGTGTACAGTGAAGAGATGGACAACTCTGTGAAGAATCTTCAAGGGTTATGG 1094
QY 1617 A 1617
Db 1095 A 1095

RESULT 12
AC021433/c
LOCUS 177556 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-688K17 map 4, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
AC021433
AC021433.3 GI:7417774
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
AUTHORS 1 (bases 1 to 177556)
TITLE Homo sapiens chromosome 4, clone RP11-688K17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177556)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrell,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severi,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo.A., Wu.X., Wyman,D., Ye.W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome

REFERENCE
AUTHORS

Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 177556)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campolano, A., Castile, A., Choquel, Y., Colangelo, M., Collins, S.,
 Collumore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Pieter, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 5, 2000 this sequence version replaced gi:6778505.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL
COMMENT

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: LS027
 Center clone name: 688_K_17

----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 170986 bases at least Q40
 Consensus quality: 174287 bases at least Q30
 Consensus quality: 175619 bases at least Q20
 Insert size: 174000; agarose-fp
 Insert size: 176456; sum-of-contigs
 Quality coverage: 5.1 in Q20 bases; agarose-fp
 Quality coverage: 5.0 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2292: contig of 2292 bp in length
 2293 2392: gap of 100 bp
 2393 4948: contig of 2556 bp in length
 4949 5048: gap of 100 bp
 5049 8358: contig of 3310 bp in length
 8359 8458: gap of 100 bp
 8459 12531: contig of 4073 bp in length
 12532 12631: gap of 100 bp
 12632 17861: contig of 5230 bp in length
 17862 17961: gap of 100 bp
 17962 28411: contig of 10450 bp in length
 28412 28511: gap of 100 bp
 28512 40270: contig of 11759 bp in length
 40271 40370: gap of 100 bp
 40371 56792: contig of 16422 bp in length

* 56793 56892: gap of 100 bp
 * 56893 80742: contig of 23850 bp in length
 * 80743 80842: gap of 100 bp
 * 80843 111092: contig of 30250 bp in length
 * 111093 111192: gap of 100 bp
 * 111193 144467: contig of 33275 bp in length
 * 144468 144567: gap of 100 bp
 * 144568 177556: contig of 32989 bp in length.

FEATURES
source

1..177556
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-688K17"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature

1..2292
 /note="assembly_fragment"
 2393..4948
 /note="assembly_fragment"
 5049..8358
 /note="assembly_fragment"
 8459..12531
 /note="assembly_fragment"
 12632..17861
 /note="assembly_fragment"
 17962..28411
 /note="assembly_fragment"
 clone_end:17
 vector_side:right

misc_feature

28512..40270
 /note="assembly_fragment"
 40371..56792
 /note="assembly_fragment"
 56893..80742
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left

misc_feature

80843..111092
 /note="assembly_fragment"
 111193..144467
 /note="assembly_fragment"
 144568..177556
 /note="assembly_fragment"

misc_feature

BASE COUNT 55171 a 32590 c 32371 g 56323 t 1101 others
 ORIGIN

Query Match 23.7%; Score 714.6; DB 2; Length 177556;
 Best Local Similarity 99.4%; Pred. No. 4.9e-183;
 Matches 717; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 897 CAGGAGTCCAGTGGAGGTATCAAGAATCTCTGGTGAACAGGTTTTTAAATATGTT 956
 Db |||||
 Db 25004 CAGGAGTCCAGTGGAGGTATCAAGAATCTCTGGTGAACAGGTTTTTAAATATGTT 24945

QY 957 ACGAGGAAGATGAATAACATCCTTGGGGTGGTGGAGGCACCCCTTAAGATTTTTTAAACA 1016
 Db |||||
 Db 24944 ACGAGGAAGATGAATAACATCCTTGGGGTGGTGGAGGCACCCCTTAAGATTTTTTAAACA 24885

QY 1017 GCTTCAGTACCTTCTGAAACAGACAGCCATTCAGAGAACAGAAAGGGGAGGC 1076
 Db |||||

Db 24884 GCTTCAGTACCTTCTGAAACAGACAGCCATTCAGAGAACAGAAAGGGGAGGC 24825
 QY 1077 TTGAGGAGCCCTCCATTCTATGCCCTGGATGAAGGAGATGATTTCTACATGTTTACTACT 1136
 Db |||||

Db 24824 TTGAGGAGCCCTCCATTCTATGCCCTGGATGAAGGAGATGATTTCTACATGTTTACTACT 24765
 QY 1137 TCTTCCCTAAGAGAACCCACCTCCCTGATTTCTCCCGGCATCATAAAGGCAGCTGCTCAGC 1196
 Db |||||

Db 24764 TCTTCCCTAAGAGAACCCACCTCCCTGATTTCTCCCGGCATCATAAAGGCAGCTGCTCAGC 24705
 QY 1197 TATTATATGAACGGAAGTGAAGTGTGTTTAATGCCCTCCCTGCTTCCATATGATATGA 1256
 |||||

24704	TATTATATGAAGCGGAAGTGGGAAGTGCCTTTAATGCCCTCCCTGCTTCCATTAATGATTGCA	24645
QY	GGAGTTTGTGAATCAGCGCCTACTTGTGTGTACTCCGTTTCACATGAAAGCACCACGCCAT	1316
Db	GGAGTTTGTGAATCAGCGCCTACTTGTGTGTACTCCGTTTCACATGAAAGCACCACGCCAT	24585
QY	CCCTGCCCCAGCAAAACCCAGTCCTCGCTGGTGATTTCCACATCGCTATTCTGCAGA	1376
Db	CCCTGCCCCAGCAAAACCCAGTCCTCGCTGGTGATTTCCACATCGCTATTCTGCAGA	24525
QY	CATTTCCATTTCCATTTTCATGTTTGACAAGATATGACAATTCGCAATTTTCGCAATTTG	1436
Db	CATTTCCATTTCCATTTTCATGTTTGACAAGATATGACAATTTTCGCAATTTTCGCAATTTG	24465
QY	TCAGAAGGCTGATGAACGAGGAGACTTTTCAGGAAAGCCTTAATTTTGAAGAACTATTG	1496
Db	TCAGAAGGCTGATGAACGAGGAGACTTTTCAGGAAAGCCTTAATTTTGAAGAACTATTG	24405
QY	AAATTCGTACTCCAAAATCAACAGAGCTTTACGGGATCATGACTATGTTGAATATGC	1556
Db	AAATTCGTACTCCAAAATCAACAGAGCTTTACGGGATCATGACTATGTTGAATATGC	24345
QY	AGTTTGTTCAGAGTGAGGAGATGGGACAACTCTGTGAAGAAATCTTCAAGGGTTATGG	1616
Db	AGTTTGTTCAGAGTGAGGAGATGGGACAACTCTGTGAAGAAATCTTCAAGGGTTATGG	24344
QY	1617 A 1617	
Db	24284 A 24284	

RESULT 13	
AC104083	196537 bp DNA linear PRI 21-FEB-2002
LOCUS	
DEFINITION	Homo sapiens BAC clone RP11-588K22 from 4, complete sequence.
ACCESSION	AC104083 AC023069
VERSION	AC104083.3 GI:18642949
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 196537)
JOURNAL	Sulston,J.P. and Waterston,R.
MEDLINE	Toward a complete human genome sequence
PUBMED	Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE	99063792
AUTHORS	9847074
TITLE	2 (bases 1 to 196537)
JOURNAL	Levy,A., Kozlowicz,A. and Boyer,E.
PUBMED	The sequence of Homo sapiens BAC clone RP11-588K22
REFERENCE	Unpublished (2001)
AUTHORS	3 (bases 1 to 196537)
TITLE	Waterston,R.H.
JOURNAL	Direct Submission
PUBMED	Submitted (03-DEC-2001) Genome Sequencing Center, Washington
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
AUTHORS	MO 63108, USA
TITLE	4 (bases 1 to 196537)
JOURNAL	Waterston,R.H.
PUBMED	Direct Submission
REFERENCE	Submitted (09-FEB-2002) Genome Sequencing Center, Washington
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
TITLE	MO 63108, USA
JOURNAL	5 (bases 1 to 196537)
PUBMED	Waterston,R.
REFERENCE	Direct Submission
AUTHORS	Submitted (21-FEB-2002) Department of Genetics, Washington
TITLE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL	On Feb 9, 2002 this sequence version replaced gi:17981676.
PUBMED	----- Genome Center.
COMMENT	Center: Washington University Genome Sequencing Center
	Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0588K22
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

The RpoI-11 human BAC library was made from the blood of one male donor as described by Osogawa, K. Wood, P. Y. Zhu, B. Freggen, E., Tateno, M., Catanesi, J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACs3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-487K5. Actual start of this clone is at base position 1 of RP11-588K22; actual end is at base position 196537 of RP11-588K22.

There are single plasmid subclone regions from 28047 to 28111 and 110730 to 110735. Data from AC021433 was used to finish this clone, AC104083.

The sequence of AC023069 has been incorporated into AC104083.

```

FEATURES
source
the sequence of reads3 has been
Location/Qualifiers
1. .196537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="rp11-588K22"
/clone_lib="RPC11"
4. .2246
/rpt_family="L1"
repeat_region
2250. .2307
/rpt_family="(TA)n"
repeat_region
2308. .2670
/rpt_family="L1"
repeat_region
2823. .3063
/rpt_family="L1"
repeat_region
3350. .3462
/rpt_family="L1"
repeat_region
3709. .4013
/rpt_family="Alu"
repeat_region
4029. .4109
/rpt_family="AT-rich"
repeat_region
4922. .5099
/rpt_family="L1"
repeat_region
5115. .5506
/rpt_family="ERV1"
repeat_region
5538. .5837

```


AB062171 3022 bp mRNA linear VRT 19-JUL-2001
 LOCUS Takifugu rubripes FrGCS-alpha mRNA for soluble guanylyl cyclase
 DEFINITION alphan subunit, complete cds.
 ACCESSION AB062171
 VERSION AB062171.1 GI:14495183
 KEYWORDS Takifugu rubripes cDNA to mRNA.
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1
 Morinaga,C., Yamamoto,T., Moriya,Y. and Suzuki,N.
 Identification of tandem organization of soluble guanylyl cyclase
 alphan and betan subunit genes in the Japanese pufferfish (Fugu
 rubripes) genome
 Unpublished
 2 (bases 1 to 3022)
 Morinaga,C., Yamamoto,T., Moriya,Y. and Suzuki,N.
 Direct Submission
 Submitted (23-MAY-2001) Norio Suzuki, Hokkaido University, Division
 of Biological Sciences, Kita 10, Nishi 8, Kita-ku, Sapporo,
 Hokkaido 060-0810, Japan (E-mail:norio-s@sci.hokudai.ac.jp,
 Tel:81-11-706-4908, Fax:81-11-746-1512)
 Location/Qualifiers
 1..3022
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /note="synonym:Fugu rubripes"
 1..3022
 /gene="FrGCS-alpha"
 226..2253
 /gene="FrGCS-alpha"
 /codon_start=1
 /product="soluble guanylyl cyclase alphan subunit"
 /protein_id="BAB60907.1"
 /db_xref="GI:14495184"
 /translation="MFCTKLKELKISGECPTSSAKTNQLEHSQLSKDVHLPSTKSD
 VRRNTGNAPHOXSRAKVNLTGIESIRKLACPEFERLHTALRRMRLSDPSREPAS
 PFVCTQYQSCDDPEHFVKWMSFSRKAIONDTLAVALEEFLNFCYREDHILRV
 VGGALHFLANSFVLLKQSLTHYQREDVCNERSVLCLODKLGLLVVYFNPHPTTE
 LFFPGVIAKARALDVTVDLMDPPATKDSILQSPSLYVQVGVVYVVDKASLSPSPL
 RSPIRGTQFQILTMNTQETIRIKRGVSTADNTLMDLQOMIVVPESDAIFLGSPC
 VKLEELTGRGLYLSLIP INHALRDVVLVGEQAKADGLKRLGKAKALEHAHAOLE
 EEKKTVDLLLETIPPGVAQQLHQGTQVQAKPERVTMLPSDIDVFTAVCSHCTPMQV
 ITMLNELYTFDHHCGLDVKVETIGDAICVAGLHKESETHQVVALMALKMWEUS
 NGVMTPAGEPIQMRIGLHTGSVLGVVGVKMPRYCLFGNNVTLANRKFESCSQPKINI
 SPVTHRLLEDHPEVFIPRRRQELPANFPEDIPGVCFLEASPRPSEVTLK"
 BASE COUNT 769 a 766 c 726 g 761 t
 ORIGIN
 Query Match 21.9%; Score 660.2; DB 5; Length 3022;
 Best Local Similarity 64.7%; Pred. No. 1.8e 168;
 Matches 1034; Conservative 0; Mismatches 553; Indels 12; Gaps 3;
 QY 909 TGGAGGTATCAAGAATCTCTGGTGAAGAGGTTTTTAAATATGTTACGAGGAAGATG 968
 DB 629 TGGTACTCTGAGTAGCGCTTGGCAGGAGCTTTTCAACATGTGTATGAAGAGACA 688
 QY 969 AAAACATCTTGGGTGGTGGAGCACCCCTTAAAGATTTTTTAAACAGCTTCAGTACCC 1028
 DB 689 GGCACATTTTGGAGGTGTGGGGGGAGCCCTGCACGACTTCTCAACAGCTTCAATGCT 748
 QY 1029 TTCTGAACAGCAGCAGCATTTGCCAAGACGAGCAAAAAGGGGAGCGCTTGAGCAGCGCT 1088
 DB 749 TGGTGAACAGCAGCAGCAGCGCTGTACCATACCAGGACAGAGAGATTGTGTAACCAACCTT 808
 QY 1089 CCATTCATGCTGGATGAAGAGGATGATTTTCTACATGTTTACTACTTCTTCCCTTAAGA 1148
 DB 809 CGGTGCTGTGTAGACAAGGATTGGTCTGCTTACTGCTATTCTTCTTCAATCCCCACC 868

1149 GAACCACTCCCTGATTCTTCCCGCATCATAAAGCAGCTGCTCAGTATATATAA 1208
 DB 869 CGACCACTGAGCTCTTTTCCCGGAGTCAACAAGCTGCTGCCGCTGCTCTATCACA 928
 QY 1209 CGGAAGTGAAGTGTCTGTTAATGCTCCCTGCTTCCATATATGATTGATGAGCAGTGTGGA 1268
 DB 929 CCACGTGGATGTTCTGTATGACCTCTGTCACATAAAGATAGCATCTCGACGACGCC 988
 QY 1269 ATCAGCCCTACTTCTGTACTCCCTTACATGAAAGACACAAG---CCATCCCTGTCCC 1325
 DB 989 CGCAGCCAGTCTTCTGTACACAGTTGTAGTAAAGATGCTTAAAGTCTTGAGCCCCAGTC 1048
 QY 1326 CCAGCAAAACCCAGTCTCTGCTGCTGATTTCCACATCGCTATTCTTCAAGACATTTCCAT 1385
 DB 1049 CACTACAGCCACCTCAGCGGGAGCTTCTTACTCTCTGTTCTCCACCATCTTCCCTT 1108
 QY 1386 TCCATTTCATGTTTGACAAAGATATGACAATCTGCAATTTGGCAATGGCAATGAGACGAC 1445
 DB 1109 TCCATCTGATCTGGACGAGCTTGTCTGTGCAAGTAGGACACGGGCTGAGGAAGA 1168
 QY 1446 TGTGAACAGGAGAGA---CTTTCAAGAAAGCCCTAATTTTGAAGAATACTTTCAAAATTC 1502
 DB 1169 GACTCACAGAAAGATGGAGTGGAGTGGACGACGACGACCTTCCAAAGACACTTCTCCATTC 1228
 QY 1503 TGACTCCAAAATCAACGACGAGCTTTAGCGGATCATGACTATCTTGAATATGACGATTG 1562
 DB 1229 TTTCTCCCTAATCAGATGATCTTCCAAGTATTCTTAACCATCTGAACACTCAGTTCA 1288
 QY 1563 TTGTACAGTGAAGAGATGGACAACCTCTGTGAAGAAATCTTCAAGGGTTATGAGACCTCA 1622
 DB 1289 CCATTCCGATCAAGCGTGGAGTCTCCACCGCAGATAACAC-----GCTCATGAGCCTTA 1342
 QY 1623 AAGCCAAATGATCTACATTTGTAATCCAGTGAATCTTGTGTTTGGGGTCAACCTGTG 1682
 DB 1343 AAGTCAAGATGTTTATGTCAGAGTCCGAGCCATTTGTTTGGGCTCGCCGTGTG 1402
 QY 1683 TGGACAGATTAGAAGATTTTACAGGACGAGGCTTCTACCTCTCAGACATCCCAATTCACA 1742
 DB 1403 TAGACAAGCTGGAGGAGCTGACGGTCTGTCGCTGTACCTGTGACATATCCCATTCATA 1462
 QY 1743 ATGCACGTGAGGATGTGGTCTTATAGGGNACAGCCGAGCTCAAGATGGCTGAGA 1802
 DB 1463 ACGGCTGCGCGAGCTGCTCTGTTGGGCAACAGGCAAGCCAGGACGCGGTGAAGA 1522
 QY 1803 AGAGCTGGGGAAGCTCAAGGCTACCTTGAGCAAGCCACCAAGCCCTGGAGGAGAGA 1862
 DB 1523 ACGGCTGGGGAAGCCCAAGCAGCCCTTGGACATGCCCATCAAGCACTGGAGGAGAGA 1582
 QY 1863 AGAAAAAGACAGTAGACCTTCTGTGCTCCATATTTCCCTGTGAGGTTGCTCAGCAGCTGT 1922
 DB 1583 AGAAGAAGCCGTAGACCTCTCTTTCACCATCTTTTCCCGGCACCGTTGCGCAGCAGCTGT 1642
 QY 1923 GGCAGGCAAGTGTGCAAGCCCAAGATGCTAGTAATGTCCACATCTCTTCTCAGACA 1982
 DB 1643 GGCAGGCCAGACGGTCCAGCCCAAGAGTTTGAAGGGTTCACAATGCTCTTCTCTGACA 1702
 QY 1983 TCGTTGGGTTCACTGCCATCTGCTCCAGTGTCTCACCGCTCGAGGTCATCACCATGTCTCA 2042
 DB 1703 TCGTGGGTTTCAAGCGGTTTGTCTCTCACTGCACCCGATGCAAGTATCACCATGTGA 1762
 QY 2043 ATGCACGTGTACACTCGCTTTCACACAGAGTGTGGAGAGTGGATGTCTACAGGTGAGA 2102
 DB 1763 ACGAGCTGTACACCAAGTTTGACCAACCTCTGGAGAGCTGGAGCTGTATAAGGTGAGA 1822
 QY 2103 CCATTGGCGATCCCTATTGTGTAGCTGGGGATTTACAAAGAGAGTGTACTCTATCTGCTG 2162
 DB 1823 CCATCGGTGATGCGTATTGCTAGCCCGGCTTTACACAGGAGGAGGAGCTCAGCGTG 1882
 QY 2163 TTCAATAGGCTGATGCCCTGAAGATGATGGAGCTCTCTGTATGAAGTATGTCTCCCC 2222
 DB 1883 TCCAAGTAGCACATCATGGCCTTAAAGATGATGGAGCTTTTCGACGCGGTGTATGACTCCAG 1942

Db 1461 GGATGGCTTGAAGAAGCGGTGGGAAAGCCAGGCGCTTGGAAGACGCTCACCAGC 1520
Qy 1849 CCTGGAGGAGAGAGAAAGACAGTAGACCTTCTGTGTCTCCATATTTCCCTGTGAGGT 1908
Db 1521 GCTGGAGGAGAGAGAGAGACAGTAGATCTCTCTTTTCCATTTTCCAGGCACCGT 1580
Qy 1909 TGCTCAGCAGCTGTGGCAAGGCAAGTTGTGAAGCCAAAGATTCAGTAATGTCCACAT 1968
Db 1581 GGCCACAGCAGCTGTGGCAGGGGAAACTGTTCAAGCAAGAAAGTTAAACCAAGTTACGAT 1640
Qy 1969 GCTCTTCTCAGACATCGTTGGGTTCACTGCCATCTGCTCCAGTGTCTCAGCGTGCAGGT 2028
Db 1641 GCTCTTCTCTGATATTGGGCTTCACAGCGGTTTGTCTCTGTGCACACCAATGCAAGT 1700
Qy 2029 CATCACCATGCTCAATGACTGTACACTGCTCGACCAGCAGTGTGGAGAGCTGGATGT 2088
Db 1701 GATCACCATGCTCAATGAGCTGTACACTAAATTTGACTATCAATGCCGAGAGCTTGATGT 1760
Qy 2089 CTACAAGTGGAGACCATTGGCGATGCTTATTTGTAGTGGGGATTACACAAAGAGAG 2148
Db 1761 CTATAAGGTAGAGACCATTAGTGATGCTTATTTGTAGCCGTGGCTTGCACAGGAGAG 1820
Qy 2149 TGATACATCTGCTCTCAGATAGCGCTGATGGCCCTGAAGATGATGAGCTCTCTGATGA 2208
Db 1821 TGAACCCATGCAAGTCGAAATTCGATTCATGGCATTTGAAGATGATTGAGCTTTGAGATGA 1880
Qy 2209 AGTTATGCTCCCCATCGAAGCCTATCAAGATGCGAATTTGGACTGCACCTCTGGATCAGT 2268
Db 1881 AGTTTTCACACCAACCGAGAGCCAAATACAGATGCGTATCGGTCTCCACAGTGGGTCACT 1940
Qy 2269 TTTTGTGGCGTGGTTGGAGTTAAATGCCCCGTTACTGTCTTTTGGAAACAATGTAC 2328
Db 1941 GTTGGCTGGTGTGTTGGGTGAGGATGCCAAGGTACTGCTGTTTGGAAACAATGTTAC 2000
Qy 2329 TCTGGCTAACAAATTTGAGTCTGCTGAGTGTACACGAAATCAATGTCAGCCCAACAC 2388
Db 2001 ATTAGCCCAACAAGTTTCAAGTCATGACGCCAGCCGAGAAAGTCAACATCAGCCCAACAC 2060
Qy 2389 TTACAGATTACTCAAGACTGTCTGCTGTTTTCGTTTACCCCTCGATCAAGGGGAGAACT 2448
Db 2061 GCACAGATTGTCAAAGGTCGCCAGAGTTTGTATTTTCCAGGAGCAGACAAAGACCT 2120
Qy 2449 TCCACCAACTTCCCTAGTAGTAATCCCGGAATCTGCCATTTTCTGGATGCTTACCAACA 2508
Db 2121 TCCGGCCAACTTCCAGAACATCCCTGCTGTTTGTACTTTTGGAGTCAGCCTTGAA 2180
Qy 2509 AGGAACAAA 2517
Db 2181 AATGTCAA 2189

Search completed: July 1, 2003, 16:02:28
Job time : 5176.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 10:46:42 ; Search time 4186.4 Seconds
(without alignments)
16983.119 Million cell updates/sec

Title: US-09-762-767A-3
Perfect score: 2443
Sequence: 1 cccccccgcgcgtgcgcg.....aaatcatttttaaaaggagt 2443

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.in.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rnd.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em_sy.*
- 39: em.htgo.hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2443	100.0	2443	6	AX024601	AX024601 Sequence
2	2443	100.0	2443	9	HSGCSAB	X66533 H.sapiens s
3	1842	75.4	1951	9	AF020340	AF020340 Homo sapi
4	1774	72.6	3137	4	BTGUCYC	Y00770 Bovine mRNA
5	1711.4	70.1	3235	10	AF297083	AF297083 Mus muscu
6	1691.8	69.3	3047	10	RATGCS	M2562 Rat solubile
7	1652.4	67.6	2303	10	AF020339	AF020339 Mus muscu
8	1151.6	47.1	2997	5	AB000850	AB000850 Oryzias l
9	1087.8	44.5	2769	5	AB062172	AB062172 Takifugu
10	582.2	23.8	2593	3	AB062386	AB062386 Hemicentr
11	559.6	22.9	3691	3	AF062751	AF062751 Manduca s
12	521	21.3	70709	9	AC114761	AC114761 Homo sapi
13	521	21.3	177556	2	AC021433	AC021433 Homo sapi
14	452.2	18.5	143255	2	AC015996	AC015996 Homo sapi
15	371.4	15.2	2840	3	DMU27123	U27123 Drosophila
16	270.6	11.1	325	4	AF486295	AF486295 Ovis arie
17	267	10.9	5575	3	AB062387S4	AB062387 Hemicentr
18	218	8.9	29480	2	AC121563	AC121563 Mus muscu
19	213.2	8.7	218756	2	AC128391	AC128391 Rattus no
20	212.4	8.7	8133	3	AF017062	AF017062 Anopheles
21	203.6	8.3	218756	2	AC128391	AC128391 Rattus no
22	203.4	8.3	2657	10	AF109963	AF109963 Rattus no
23	200.2	8.2	2954	6	AX281673	AX281673 Sequence
24	200.2	8.2	2954	9	HSGCA2	X63282 H.Sapiens m
25	192.6	7.9	2258	9	HSU58855	U58855 Homo sapien
26	192.6	7.9	2627	9	AK096588	AK096588 Homo sapi
27	192.6	7.9	2657	9	BC028384	BC028384 Homo sapi
28	192.6	7.9	3015	6	AX024599	AX024599 Sequence
29	192.6	7.9	3015	9	HSRNASGC	Y15723 Homo sapien
30	187.2	7.7	2445	10	AF297082	AF297082 Mus muscu
31	186.2	7.6	3004	9	HSGCSAA	X66534 H.sapiens s
32	178.2	7.3	2668	10	RATGCSB	M57405 Rat solubile
33	178.2	7.3	4775	10	RNU60835	U60835 Rattus norv
34	177	7.2	3261	4	BTGUCV73	X54014 Bovine mRNA
35	174.2	7.1	3022	5	AB062171	AB062171 Takifugu
36	167.4	6.9	9762	5	AB062170	AB062170 Takifugu
37	167.4	6.9	17366	5	AB022281	AB022281 Oryzias l
38	160.6	6.6	3047	9	HSRNASCA2	Z50053 H.sapiens m
39	150.4	6.2	2570	5	AB000849	AB000849 Oryzias l
40	149.4	6.1	62894	2	AC113467	AC113467 Mus muscu
41	147.8	6.0	661	10	AB079780	AB079780 Rattus no
42	147.8	6.0	206456	2	AC112402	AC112402 Rattus no
43	147.2	6.0	2335	6	AX401893	AX401893 Sequence
44	147.2	6.0	2335	10	RATGCSB2	M57507 Rattus norv
45	147.2	6.0	2452	10	AY004153	AY004153 Rattus no

ALIGNMENTS

RESULT 1
AX024601
LOCUS AX024601 Sequence 3 from Patent DE19837015.
DEFINITION Sequence 3 from Patent DE19837015.
ACCESSION AX024601
VERSION AX024601.1 GI:10184740
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2443)
AUTHORS Poller.W., Schmidt,H. and Zabel,U.
TITLE New human soluble guanylate cyclase alpha/beta1 and the nucleic acid encoding the subunits, useful for producing diagnostic

antibodies, and for somatic gene therapy of arteriosclerosis									
Patent: DE 19837015-A 3 24-FEB-2000;									
VASOPHARM BIOTECH GMBH & CO KG (DE)									
FEATURES	Location/Qualifiers								
	1..2443								
BASE COUNT	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
ORIGIN	709 a	513 c	533 g	688 t					
Query Match 100.0%; Score 2443; DB 6; Length 2443;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 2443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	CCCCCCCCCGCGCTGCCGCTCTGCTGGGTCCCTTCGGCCGTACCTCTGCGCTGGGGC	60						
Db	1	CCCCCCCCCGCGCTGCCGCTCTGCTGGGTCCCTTCGGCCGTACCTCTGCGCTGGGGC	60						
QY	61	TGCTCCCGCGCTCCCGGTGCAGACACCATGTACGGATTTGTGAATCAGCCCTCGAGTT	120						
Db	61	TGCTTCGCCGCTCCCGGTGCAGACACCATGTACGGATTTGTGAATCAGCCCTCGAGTT	120						
QY	121	GCTGTGATTCGCAATTCACGCCCGCGAGTGTGGGAAGACATCAAAAAGAGGCACAGTT	180						
Db	121	GCTGTGATTCGCAATTCACGCCCGCGAGTGTGGGAAGACATCAAAAAGAGGCACAGTT	180						
QY	181	AGATGAAGAGGACAGTTTCTGTCAGAAATATATATGATGACTCCAAAACCTTATGATTT	240						
Db	181	AGATGAAGAGGACAGTTTCTGTGCAAGTAATATATGATGACTCCAAAACCTTATGATTT	240						
QY	241	GGTGTCTGTCGAAGCAAGTCTCAATCTCAATGCTGGAGAAATCCTCCAAATGTTTG	300						
Db	241	GGTGTCTGTCGAAGCAAGTCTCAATCTCAATGCTGGAGAAATCCTCCAAATGTTTG	300						
QY	301	GAAGATGTTTTCGTCCTTTGCCAAGAAATCTGGTTATGATACAACTCTGGTGCTCTGG	360						
Db	301	GAAGATGTTTTCGTCCTTTGCCAAGAAATCTGGTTATGATACAACTCTGGTGCTCTGG	360						
QY	361	CTCTAATGTGAGAGAAATTTCTACAGAACCTTGATGCTCTGCAGGACACCTTGCTACCAT	420						
Db	361	CTCTAATGTGAGAGAAATTTCTACAGAACCTTGATGCTCTGCAGGACACCTTGCTACCAT	420						
QY	421	CTACCCAGGAATGCGTGACCTTCTTTAGTGCACTGATGAGAAAAGGCGAAGGACT	480						
Db	421	CTACCCAGGAATGCGTGACCTTCTTTAGTGCACTGATGAGAAAAGGCGAAGGACT	480						
QY	481	CATTTTGCCTACTACTCAGAGAGAGGACTTTCAGGATATTGTCATTGGAATCATCAA	540						
Db	481	CATTTTGCCTACTACTCAGAGAGAGGACTTTCAGGATATTGTCATTGGAATCATCAA	540						
QY	541	AACAGTGGCACAACAAATCCATGGCACTGAAATAGACATGAAGTTATTTCAGCAAGAA	600						
Db	541	AACAGTGGCACAACAAATCCATGGCACTGAAATAGACATGAAGTTATTTCAGCAAGAA	600						
QY	601	TGAAGAATGTGATCATCAATTTTAAATGGAAGAAAAGAGCTCAAAAGAGAGGATTT	660						
Db	601	TGAAGAATGTGATCATCAATTTTAAATGGAAGAAAAGAGCTCAAAAGAGAGGATTT	660						
QY	661	TTATGAAGATCTTGACAGATTTTGAAGAAAATGGTACCAGGAATCACGCATCAGCCATA	720						
Db	661	TTATGAAGATCTTGACAGATTTTGAAGAAAATGGTACCAGGAATCACGCATCAGCCATA	720						
QY	721	TACATTCGCAAGCTTTTCCCTTTTCATATATATTTGACCGGGACCTAGTGGTCACTCA	780						
Db	721	TACATTCGCAAGCTTTTCCCTTTTCATATATATTTGACCGGGACCTAGTGGTCACTCA	780						
QY	781	GTGTGGCAATGCTATATACAGAGTTCTCCCGCAGCTCCAGCCTGGGAATTCAGCCTTCT	840						
Db	781	GTGTGGCAATGCTATATACAGAGTTCTCCCGCAGCTCCAGCCTGGGAATTCAGCCTTCT	840						
QY	841	GTCTGTCTTCGCTGGTTCCTCTCATATTGATATTAGTTTCCATGGGATCCTTTCTCA	900						
Db	841	GTCTGTCTTCGCTGGTTCCTCTCATATTGATATTAGTTTCCATGGGATCCTTTCTCA	900						

QY	901	CATCAATACCTGTTTTTGTATTGAGAAAGCAAGAGGATTGTTGGATGTGGAGAAATTAGA	960
DB	901	CATCAATACCTGTTTTTGTATTGAGAAAGCAAGAGGATTGTTGGATGTGGAGAAATTAGA	960
QY	961	ATGTGAGGATGAACCTGACTGGGACTGAGATCAGCTCTTACGTTCTCAAGGGTCAAAATGAT	1020
DB	961	ATGTGAGGATGAACCTGACTGGGACTGAGATCAGCTCTTACGTTCTCAAGGGTCAAAATGAT	1020
QY	1021	CTACTTACCTGAAGCAGATAGCATACTTTTCTATGTTTCAACCAAGTGTCTATGACCTGGA	1080
DB	1021	CTACTTACCTGAAGCAGATAGCATACTTTTCTATGTTTCAACCAAGTGTCTATGACCTGGA	1080
QY	1081	CGATTTGACAAGGAGAGGCTGTATCTAAGTGCATCCCTCTGCTGATGATGCCACGGCGCA	1140
DB	1081	CGATTTGACAAGGAGAGGCTGTATCTAAGTGCATCCCTCTGCTGATGATGCCACGGCGCA	1140
QY	1141	TCCTGTTCTTTTGGGAGAACAAATTTAGAGAGGAATACAACTCACCNAAGACTGGAAT	1200
DB	1141	TCCTGTTCTTTTGGGAGAACAAATTTAGAGAGGAATACAACTCACCNAAGACTGGAAT	1200
QY	1201	CCTCACTGACAGCTACAGCTCACGTTAAGAGCCCTGGAGATGAAAAGAAAAGACAGA	1260
DB	1201	CCTCACTGACAGCTACAGCTCACGTTAAGAGCCCTGGAGATGAAAAGAAAAGACAGA	1260
QY	1261	CACATTGCTGTATTCTCTCCGCTCTGTGCAATGAGCTGCGGCACAAAGCGTCC	1320
DB	1261	CACATTGCTGTATTCTCTCCGCTCTGTGCAATGAGCTGCGGCACAAAGCGTCC	1320
QY	1321	AGTGCCTGCCAAAGATATGACAAATGTGACCACTCTCTTTAGTGGCAATGTGGGCTCAA	1380
DB	1321	AGTGCCTGCCAAAGATATGACAAATGTGACCACTCTCTTTAGTGGCAATGTGGGCTCAA	1380
QY	1381	TGCTTTCTGTACAGAGCTGATCTGGAGAGGAGCCATGAAGATGCTCAACCTCTCAA	1440
DB	1381	TGCTTTCTGTACAGAGCTGATCTGGAGAGGAGCCATGAAGATGCTCAACCTCTCAA	1440
QY	1441	CGACCTCTACACAGATTTGACACACTGACTGATTCGCCGAAAAACCCATTTGTTTATAA	1500
DB	1441	CGACCTCTACACAGATTTGACACACTGACTGATTCGCCGAAAAACCCATTTGTTTATAA	1500
QY	1501	GGTGGAGCTGTTGGTGACAAATATGACAGTGTGTTTACAGAGCCATGCAATCA	1560
DB	1501	GGTGGAGCTGTTGGTGACAAATATGACAGTGTGTTTACAGAGCCATGCAATCA	1560
QY	1561	CCATGCACGATCCATCTGCCACTGGCCCTTGACATGATGAAATTTGCTGGCCAGGTTCA	1620
DB	1561	CCATGCACGATCCATCTGCCACTGGCCCTTGACATGATGAAATTTGCTGGCCAGGTTCA	1620
QY	1621	AGTAGATGTGAATCTGTTTCAGATAACAAATAGGGATACACACTGGAGAGGTAGTTACAG	1680
DB	1621	AGTAGATGTGAATCTGTTTCAGATAACAAATAGGGATACACACTGGAGAGGTAGTTACAG	1680
QY	1681	TGTCATAGCAGACGGATGCTGCTACTGCTTTTGGGAATACTGTCAACCTCAACAG	1740
DB	1681	TGTCATAGCAGACGGATGCTGCTACTGCTTTTGGGAATACTGTCAACCTCAACAG	1740
QY	1741	CCGAAACAGAAACACAGGAGAAAAGGAAAAATAATGTCTGCTGAATATACATACAGATG	1800
DB	1741	CCGAAACAGAAACACAGGAGAAAAGGAAAAATAATGTCTGCTGAATATACATACAGATG	1800
QY	1801	TCCTTAGTCTCCAGAAAATTCAGATCCCAANTTCCACTTTGGAGCAGAGGCCCGGCTGTC	1860
DB	1801	TCCTTAGTCTCCAGAAAATTCAGATCCCAANTTCCACTTTGGAGCAGAGGCCCGGCTGTC	1860
QY	1861	CATGAGGGCAAAAAGAGACCAATGCAAGTTTGGTTCTCTATCCAGAAAAAATACAGGAAC	1920
DB	1861	CATGAGGGCAAAAAGAGACCAATGCAAGTTTGGTTCTCTATCCAGAAAAAATACAGGAAC	1920
QY	1921	AGAGAAAACAAAGCAGGATGATGATCTTGGATTTGGGTTGAGAGGAGTACACA	1980
DB	1921	AGAGAAAACAAAGCAGGATGATGATCTTGGATTTGGGTTGAGAGGAGTACACA	1980

```
QY 1981 CTAGTTCACGTTTCTCCTAACACGTCGCCAAGCCAGGACGAGTCTCTCCCTATGGATA 2040
Db 1981 CTAGTTCACGTTTCTCCTAACACGTCGCCAAGCCAGGACGAGTCTCTCCCTATGGATA 2040
QY 2041 CAGATTTTCTTTTCTCCTTGTCCTTACCCCAAGACTTCTCTAGATATATCTCTCACT 2100
Db 2041 CAGATTTTCTTTTCTCCTTGTCCTTACCCCAAGACTTCTCTAGATATATCTCTCACT 2100
QY 2101 ATCCGTTATTCACACTTACGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160
Db 2101 ATCCGTTATTCACACTTACGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160
QY 2161 AAGTTTGGCTTTTCTGATGGATGATGAGCTTCTGATGCTTAAATCTACTCAAGCA 2220
Db 2161 AAGTTTGGCTTTTCTGATGGATGATGAGCTTCTGATGCTTAAATCTACTCAAGCA 2220
QY 2221 TTACTTAACATGGTATCTGCAAGTAGTAGGACCCCAATAATATTTTGAATTTAGTT 2280
Db 2221 TTACTTAACATGGTATCTGCAAGTAGTAGGACCCCAATAATATTTTGAATTTAGTT 2280
QY 2281 AAATGAACACTGAACAGTGTGGCCATGTGTATATTTATATCATGTTTACCAATCTGTT 2340
Db 2281 AAATGAACACTGAACAGTGTGGCCATGTGTATATTTATATCATGTTTACCAATCTGTT 2340
QY 2341 TAGTGTCCACATATATATATATATATATATATATATATATATATATATATATATAT 2400
Db 2341 TAGTGTCCACATATATATATATATATATATATATATATATATATATATATATATAT 2400
QY 2401 CATGTTGGTGTATATATATATATATATATATATATATATATATATATATATATAT 2443
Db 2401 CATGTTGGTGTATATATATATATATATATATATATATATATATATATATATATAT 2443

RESULT 2
HSGCSAB 2443 bp mRNA linear PRI 19-NOV-1992
LOCUS H.sapiens soluble guanylate cyclase small subunit mRNA.
DEFINITION X6533
ACCESSION X6533
VERSION X6533.1 GI:31685
KEYWORDS cytoplasmic protein; GTP pyrophosphate-lyase; Guanylate cyclase.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2443)
AUTHORS Guilla, G., Scholl, U., Bulle, F. and Guellaen, G.
TITLE Molecular cloning of the cDNAs coding for the two subunits of
soluble guanylyl cyclase from human brain
JOURNAL FEBS Lett. 304 (1), 83-88 (1992)
MEDLINE 92316204
PUBMED 1352257
REFERENCE 2 (bases 1 to 2443)
AUTHORS Guellaen, G.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-1992) Georges Guellaen, Unite INSERM 99, Hopital
Henri Mondor, 94010 Creteil, FRANCE
FEATURES
source location/Qualifiers
1. .2443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q31.3-q33"
/tissue_type="brain, frontal lobe"
/clone_lib="lambda Gr10"
89. .1948
/EC_number="4.6.1.2"
/note="soluble"
/codon_start=1
/product="guanylate cyclase"
/protein_id="CAA47144.1"
/db_xref="GI:31686"
/db_xref="SWISS-PROT:O02153"
/translation="MYGFVNHALLLIRVNGPEVWEDIKKEAQIDBEGQFLVRIYD
```

```
DSKYDLVAASKVLNLNAGBILQFMKMFVFCQESGYDILRLVLSNNREFQLND
ALHDHATIPGMRAFSRCDTAEKGLILHYSEREGLODIYIGIKTVAQOIHGT
EIDMKVIOQRNECDHTQFLIEEKSEDEYEDLDRFENGTOISPTPTFCFKAPF
PHILFDRLVYTCGNALRYLPOLPGNCSLLSVSLVRPHIDISPHGILSHLNTVF
VLRSKGLDVKELECEDELNGTETSLRLKAGOMLYLPEADSLIFLSPSPVNLDDLT
RRLGLYSDIPHLHATRDLYLLGEOPREYKLTLELEILDRQLTLRLALEDEKKTDT
LLSYLSPSVANELRHRKRPVPAKRYVDNVTILFSGIVGFNAFCSKHASGEAMKIVNLL
NDLYTRFDLTDSRKNPFYKVTVDGDMVTGVSGLPEPFIHARSICHIALDMMEIAG
QYVDGESVQITIGITHTGEVTVIGQPRYICLFGNTVLTSTETGEGKINVSE
YTYRCLMSPENDPQFHLHGRGPMKKEPMQVWFLSRKNTGTETTKQDDD"
BASE COUNT 709 a 513 c 533 g 688 t
ORIGIN
Query Match 100.0%; Score 2443; DB 9; Length 2443;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCCCCCCCCGCGTCCCGCTCGCCTGGCTGGCTCCCTTCGGCCGTACCTCTCGTGGGGGC 60
Db 1 CCCCCCCCCCGCGTCCCGCTCGCCTGGCTGGCTCCCTTCGGCCGTACCTCTCGTGGGGGC 60
QY 61 TGCCTCCCGCGTCCCGCTCGCCTGGCTGGCTCCCTTCGGCCGTACCTCTCGTGGGGGC 120
Db 61 TGCCTCCCGCGTCCCGCTCGCCTGGCTGGCTCCCTTCGGCCGTACCTCTCGTGGGGGC 120
QY 121 GCTGTGATCGCAATTAACGCCCGGAGGTGTGGAGACATCAAAAAGAGGACAGTT 180
Db 121 GCTGTGATCGCAATTAACGCCCGGAGGTGTGGAGACATCAAAAAGAGGACAGTT 180
QY 181 AGATGAAGAAGACAGTTTCTTGTCAGATAATATATATATATATATATATATATATATAT 240
Db 181 AGATGAAGAAGACAGTTTCTTGTCAGATAATATATATATATATATATATATATATATAT 240
QY 241 GTTTGTCTGCTCAAGCAAGTCTCAATCTCAATCTGAGAGAAATCTTCCAAATCTTTGG 300
Db 241 GTTTGTCTGCTCAAGCAAGTCTCAATCTCAATCTGAGAGAAATCTTCCAAATCTTTGG 300
QY 301 GAAGATGTTTTTCTGTTTTGCCAAGATCTGGTTATGATACAATCTTTCGGTGTCTCTGG 360
Db 301 GAAGATGTTTTTCTGTTTTGCCAAGATCTGGTTATGATACAATCTTTCGGTGTCTCTGG 360
QY 361 CTCCTAATCTCAGAGATTTCTACAGAACCTTGATGCTCTGCAGCACCACCTTGCTACCAT 420
Db 361 CTCCTAATCTCAGAGATTTCTACAGAACCTTGATGCTCTGCAGCACCACCTTGCTACCAT 420
QY 421 CTACCCAGGAATGCGTGACCTCTCTTTAGTGTGACTGATGAGAGAAAGGGCAAGAGACT 480
Db 421 CTACCCAGGAATGCGTGACCTCTCTTTAGTGTGACTGATGAGAGAAAGGGCAAGAGACT 480
QY 481 CATTTTGCACCTACTACTCAGAGAGAGAGGACTTCAGGATATTTGTCATTTGGAATCATCA 540
Db 481 CATTTTGCACCTACTACTCAGAGAGAGAGGACTTCAGGATATTTGTCATTTGGAATCATCA 540
QY 541 AACAGTGCACAAACAAATCCATGGACCTGAATAGACATGAGGTTATTCAGCAAGAA 600
Db 541 AACAGTGCACAAACAAATCCATGGACCTGAATAGACATGAGGTTATTCAGCAAGAA 600
QY 601 TCAAGAATGTGATCATACTCAATTTTAAATTAAGAGAAAGAGTCAAAAGAGAGAGATTT 660
Db 601 TCAAGAATGTGATCATACTCAATTTTAAATTAAGAGAAAGAGTCAAAAGAGAGAGATTT 660
QY 661 TTATGAAGATCTTGACAGATTTGAAGAAATGGTACCAGGAATCAGCATCAGCCCAT 720
Db 661 TTATGAAGATCTTGACAGATTTGAAGAAATGGTACCAGGAATCAGCATCAGCCCAT 720
QY 721 TACATCTGCAAGGCTTTTCTTTCATATATATATATATATATATATATATATATATATAT 780
Db 721 TACATCTGCAAGGCTTTTCTTTCATATATATATATATATATATATATATATATATATAT 780
QY 781 GTGTGGCAATGCTATATACAGAGTTCTCCCCAGCTCCAGCTCGGAAATTCAGAGCTTCT 840
Db 781 GTGTGGCAATGCTATATACAGAGTTCTCCCCAGCTCCAGCTCGGAAATTCAGAGCTTCT 840
```


Qy	841	GTCTGCTCTTCGCTGCTGCTCTCTCTCATATGATATATAGTTTCCAGGGGATCCTTTCTCA	900
Db	841	GTCTGCTCTTCGCTGCTGCTCTCTCATATGATATATAGTTTCCATGGGATCCTTTCTCA	900
Qy	901	CATCAATACATGTTTTCTTATGTGAAAGCAAGGAGGATTTGTTGGATGTGAGAAATTAGA	960
Db	901	CATCAATACATGTTTTCTTATGTGAAAGCAAGGAGGATTTGTTGGATGTGAGAAATTAGA	960
Qy	961	ATGTGAGGATGAATGACTGGGACTGAGATCAGCTGCTTACGTCTCAAGGGTCAAAATGAT	1020
Db	961	ATGTGAGGATGAATGACTGGGACTGAGATCAGCTGCTTACGTCTCAAGGGTCAAAATGAT	1020
Qy	1021	CTACTTACTGAAGCAGATAGCATACATTTTCTATGTTCCACCAGTGTCAATGAACCTGGA	1080
Db	1021	CTACTTACTGAAGCAGATAGCATACATTTTCTATGTTCCACCAGTGTCAATGAACCTGGA	1080
Qy	1081	CGATTTGACAGGAGAGGGCTGTATCTAACTGACATCCCTCTCGATGATGCCACGGCGGA	1140
Db	1081	CGATTTGACAGGAGAGGGCTGTATCTAACTGACATCCCTCTCGATGATGCCACGGCGGA	1140
Qy	1141	TCCTTGTCTTTTGGGAGAACAAATTTACAGAGGAATACAAACTCACCACAAACTGGGAAAT	1200
Db	1141	TCCTTGTCTTTTGGGAGAACAAATTTACAGAGGAATACAAACTCACCACAAACTGGGAAAT	1200
Qy	1201	CCTCACTGACAGGCTACAGTCAAGTTTAAAGAGCCCTGGAAGATGAAAGAAAAAGACAGA	1260
Db	1201	CCTCACTGACAGGCTACAGTCAAGTTTAAAGAGCCCTGGAAGATGAAAGAAAAAGACAGA	1260
Qy	1261	CACATTTGCTGTATTTCTCTCTTCCCTCGTCTGTTGCCAATGAGTGGGGCACAAAGGCTCC	1320
Db	1261	CACATTTGCTGTATTTCTCTCTTCCCTCGTCTGTTGCCAATGAGTGGGGCACAAAGGCTCC	1320
Qy	1321	AGTGCCTGCCAAAGATATGACAAATGTGACATCCTCTTTTGTGGCAATTTGGGGCTTCAA	1380
Db	1321	AGTGCCTGCCAAAGATATGACAAATGTGACATCCTCTTTTGTGGCAATTTGGGGCTTCAA	1380
Qy	1381	TGCTTTCTGTAGCAAGCATGCACTGTGAGAAAGAGCCATGAAGATGCTCAACCTCTCAA	1440
Db	1381	TGCTTTCTGTAGCAAGCATGCACTGTGAGAAAGAGCCATGAAGATGCTCAACCTCTCAA	1440
Qy	1441	CGACCTCTACACAGATTTGACACACTGACTGATCCCGGAAAAACCCATTTCTTTATAA	1500
Db	1441	CGACCTCTACACAGATTTGACACACTGACTGATCCCGGAAAAACCCATTTCTTTATAA	1500
Qy	1501	GGTGGAGACTGTTGGTGACAAAGTATATGACAGTCAAGTGGTTTACCAGAGCCATGCAATCA	1560
Db	1501	GGTGGAGACTGTTGGTGACAAAGTATATGACAGTCAAGTGGTTTACCAGAGCCATGCAATCA	1560
Qy	1561	CCATGACAGATCCATCTGCCACCTGGCCCTTTGGACATGATGAGGAAATGCTGGCCAGGTTCA	1620
Db	1561	CCATGACAGATCCATCTGCCACCTGGCCCTTTGGACATGATGAGGAAATGCTGGCCAGGTTCA	1620
Qy	1621	AGTAGATGGTGAATCTCTTCAGATACAAATAGGGATACACACTGGAGAGGTAGTTACAGG	1680
Db	1621	AGTAGATGGTGAATCTCTTCAGATACAAATAGGGATACACACTGGAGAGGTAGTTACAGG	1680
Qy	1681	TGTCATAGGACAGCGGATGCTCGATCTCTCTTTTGGGAAATGCTCAACCTCACAAG	1740
Db	1681	TGTCATAGGACAGCGGATGCTCGATCTCTCTCTTTTGGGAAATGCTCAACCTCACAAG	1740
Qy	1741	CCGAAACAGAAACACAGGAGAAAGGAAAAATAAATGTGCTGTAATATACATACAGATG	1800
Db	1741	CCGAAACAGAAACACAGGAGAAAGGAAAAATAAATGTGCTGTAATATACATACAGATG	1800
Qy	1801	TCATTATGCTCCAGAAAAATTCAGATCCACAAATTCACATTTGGGACACAGAGGCCAGTGTC	1860
Db	1801	TCATTATGCTCCAGAAAAATTCAGATCCACAAATTCACATTTGGGACACAGAGGCCAGTGTC	1860
Qy	1861	CATGAAGGGCAAAAAAGAACCAATGCAAGTTTCTGTTTCTATCCAGAAAAAATACAGGAAC	1920
Db	1861	CATGAAGGGCAAAAAAGAACCAATGCAAGTTTCTGTTTCTATCCAGAAAAAATACAGGAAC	1920
Qy	1921	AGAGGAACAAAGCAGGATGATGACTGAATCTTTGGATTTATGGGGTGGAGAGGAGTACAGA	1980

Qy	2163	GTTT-GCCTTTTGATGGGAAGTAGTGAGCTTCATGTGCTCTAAAATCTACTACAAGCAT	2222
Db	2190	TTTTTGGGTTTTTGCATGGTGAAGTCTGAGC- ---TGTTCCTTAATTTCTCTACAAGCAT	2245
Qy	2222	TACCTAACATGGTGATCTGCAAGTAGTAGGCCACCCTAAT---AAATATTGTTGTAATTTAG	2278
Db	2246	TGCGCTAGACGGAGCGTCTCAAAGTGATAAGCACCCACTGRTGTAAGTTGTTTAAATCTGA	2305
Qy	2279	TTAAATGAAATGAACAGAGTGTGGGCATGTGTATATTTATATATCATGTTTACCACAAATCTG	2338
Db	2306	TAGAAGCAGACTTAAYTAGTATCTGGCCATCGGTATATAT-CATGGCTCAGTAGATTG	2364
Qy	2339	TTTAGTGTGCCACATATA-----TGTATATGATATATTTTAAAGTACATA--ATGTAATA	2390
Db	2365	TTTTATGCTCCATGTATATGCTGTGTATATGATATATTTTAAAGTACATAACCAAAAA	2424
Qy	2391	AAGTTTATATCATGTGGTGTATATCATATATAGAAATCATTTCTTAAGAGT	2443
Db	2425	AAGTTTATATCATGTGGTGCATGGCATCTTAGAACCAATTTGTACAGAGT	2477
RESULT	6		
RATGS			
LOCUS		3047 bp mRNA linear ROD 27-APR-1993	
DEFINITION		Rat soluble guanylate cyclase 70kd subunit mRNA, complete cds.	
ACCESSION	M22562		
VERSION	M22562.1	GI:204273	
KEYWORDS		guanylate cyclase.	
SOURCE		Rat (Wistar; adult) lung, cDNA to mRNA, clones 70-36, 36-8, 36-16.	
ORGANISM		Rattus norvegicus	
REFERENCE		1. (bases 1 to 3047)	
AUTHORS		Nakane,M., Saheki,S., Kuno,T., Ishii,K. and Murad,F.	
TITLE		Molecular cloning of a cDNA coding for 70 kilodalton subunit of	
JOURNAL		soluble guanylate cyclase from rat lung	
MEDLINE	Biochem. Biophys. Res. Commun. 157 (3), 1139-1147 (1988)		
PUBMED	89087429		
COMMENT		Draft entry and computer-readable sequence for [1] kindly provided by M.Nakane, 15-FEB-1989.	
FEATURES			
source		Location/Qualifiers	
mRNA		1..3047	
CDS		/organism="Rattus norvegicus"	
		/db_xref="taxon:10116"	
		<1..3047	
		/product="SGC mRNA"	
		42..1901	
		/notes="guanylate cyclase 70kd subunit (BC 4.6.1.2)"	
		/codon_start=1	
		/protein_id="AAA41204.1"	
		/db_xref="GI:204274"	
		/translation="MYGFVNHAELELVIRNYPGEVVEDIKKALDQEGQLRVRIYIDHSTYDLVAASXVLNLNAGEIILQMFGKMFVFQCSEGYDTILRVLGSNNREFQLNDALHDLYLATVPGRAPSFRCDADEKGGLTLHYYSERGEVDIGYIKITVAQIHGT EIDMKYIQRSCECHDTFLIEESKEDEDFEFDLRFENGQDSRIISPYTFCKAFPI FHIIFDRLLVTOCGNAIVLPQKCSLSVSFLVRPHDISFPHGLSLHNIVPFL RSKSLDLVEKLEDGETGAELSCRLRKGMXILPEADSIILFCSPSYMNLDDLT REGJLSVDPHLNATRDVLVGQFRFEETKLDTLEILLTDLQTLAKADEKKIVT LLYSVLPSPVANERHKRPFAKRYDNVTILFSGIFNAFCSHKASGEGAMKVALL QNDYLFDTLDTSRKPFPVKYTVGDGVKMTVSLGPCPIHHARSHICHALDMMEVTA QVOVDGESVOITGITGTEGVGVIGORMPRYLFGNTVNLTSGTSTETTTPGKGINVSE YTRCLMSPENSDQPHELRHPGPSVMGKKEPPMQVFLSRKNGTTEETNQDEN"	
BASE COUNT	850 a 656 c 691 g 850 t		
ORIGIN		244 bp upstream of psti site.	
Query Match		69.3%; Score 1691.8; DB 10; Length 3047;	
Best Local Similarity		83.6%; Pred. No. 0;	
Matches 2019,		Conservative 0; Mismatches 372; Indels 24; Gaps 8;	
Qy	50	TGCGTGGGGCTGCCTCCC CGGTCCCGGTGCAGACACCATGATGACGATTGTGAATCAC	109

32	CGCGCGTGTAGTCGCTGCTGGGTCCCTTCCGCGGTACTCTCTGGCCGTGAAGCTGCTTCCT	91
69	CGGCTCCCGGTGCAGACACCACTATGAGGATTTGTGAATCACGCCCTGGAGTTGCTGGTGA	128
92	AGGCTCCGGCGCGGACACCACTGTACGGTTTCGTGAACCACTGCCCTGGAGCTGCTGGTGA	151
129	TCCGCCAATTACGCCCCCGAGGTGTGGGAAGACATCAAAAAGAGGCACAGTTAGATGAAG	188
152	TCCGCCAATTATGGTCCCGAGGTGTGGGAAGACATCAAAAAGAGGCACAGCTGGATGAAG	211
189	AAGGACAGTTCTTGTGCAGATAAATATATGATGACTCCAAAACCTATGATTTGGTTGGT	248
212	AAGCCAGTTTCTGTGCAGATAAATATACGATGATTCAAAACATATGACCTGGTGCCCTG	271
249	CTGCAAGCAAGTCCCTCAATCTCAATGCTGGAGAAATCCTCAAATGTTTGGGAAGATGT	308
272	CTGCAAGCAAGTCCCTCAACCTCAATGCTGGCGAAATCTGCAGATGTTTGGGAAGATGT	331
309	TTTTTCGTCCTTTGCCAAGAACTGGTTATGATACAAATCTTGGTGTCTGGGCTCTAATG	368
332	TTTTTCGTCCTTCTGCCAGGAGTCTGGCTATGATACCAATCTTGGGTGTCTGGATCTAATG	391
369	TCAGAGAAATTTCTACAGAAACCTTTGATGCTCTGCACGACCACCTTGCTACCACTTACCCAG	428
392	TCAGAGGTTTTTGCAGAACCTCGATGGCCCTGATGACCACCTCGCCACCATTTACCCAG	451
429	GAATGGTGCACCTTCCTTTAGGTGCACCTGATGCAGAAAGGCAAGGACTCATTTTGC	488
452	GGATGCGCGGCTTCCTTTCAGGTGCACCGATCGCGAGAAAGCAAGGCTCATCTCTC	511
489	ACTACTCTCAGAGAGAGAGGACTTTCAGGATATTGTCATTTGGAATCATCAAAACAGTGG	548
512	ACTPACTACTCGGAAGAGAGGGGCTTCAGGACATCTGATCGGATATATCAAGACTGTTG	571
549	CACAACAAATCCATGCCACTGAAATAGACATGAAGCTTATTCAGCAAGAAATGAAGAT	608
572	CTCAACAGATACACGGCACTGAGATGAGCATGAAGTGTGATTCAGCAAGAAATGAAGAT	631
609	GTGATCATATCTCAATTTTTTAATTGAAGAAAGAGTCAAAAGAAAGAGATTTTATGAAG	668
632	GTGATCACACCCAGTTTTTAATTGAAGAAAGAAATCAAAAGAGAGATTTTATGAAG	691
669	ATCTTGCAGAGATTTGAAGAAATGGTACCAGGAATTCAGCATCAGCCCATATACATTTCT	728
692	ATCTGTATAGTTTGAAGAGAAATGTTACCCAGGAATCAGGTATCAGCCCTTACACCTTCT	751
729	GCAAGCTTTTCCTTTTTCATATAATATTGACCGGGACCTAGTGTGTCATCAGTGTGGCA	788
752	GCAAGCATTTCTCTTTTCATATATATTGACCGGAACCTAGTGTGTCATCAGTGTGGCA	811
789	ATGCTATATACAGAGTTCTCCCCAGCTCCAGCTGGGAATTCAGCCCTCTGCTGTCT	848
812	ATGCCATCTACAGTGTCTCCCCAGCTCCAGCTGGGAATTCAGCCCTCTGCTGTCT	871
849	TCTCGCTGGTTCCTCTCATATTGATATATAGTTTCCATGGGATCCCTTCTCACAATCAATA	908
872	TCTCTCTGGTCCGCCCTTCATCGACATCAGTTTCCATGGGATTTCTTTCACACATCAATA	931
909	CTGCTTTTGTGATTGAGAACCAAGGAGGATTTGTTGATGTGGAGAAATTAGAATGTGAGG	968
932	CAGTCTTTGTGACTGGAAGCAAGGAGGTTGCTGTGATGTGGAGAACTTGATGTGTGAGG	991
969	ATGAACCTGACTGGGACTGAGATCAGCTGCTTACGTCTCAAGGCTCAAAATGATCTACTTAC	1028
992	ATGACTGACTGGAGCAGAGATTTAGTGTCTTACCTCTCAAGGCCAAATGATCTACTTAC	1051
1029	CTGAAGCAGATAGCATACTTTTTCTATGTTTCAACAAGTGTCAATGACCTGGAGATTTGA	1088
1052	CAGAAGCAGATAGCATCTCTCTCTCTGTTTCAACAAGTGTGATGAACCTGGACGACCTAA	1111
1089	CAGGAGAGGGCTGTATCTAAGTGACATCCCTCTGATGATGATCCAGCGGATCTTTGTTT	1148
1112	CAGAAGAGCCCTGTATCTCAGTGAGATCCCTCTCCACGATGCTACCGAGACTGGTTC	1171

Db 311 AAATGGGAAGAAATCAAAACGGCAAGCCGAGTGGAGATCGAAGGAGTTTCTTAGTCGGA 370
QY 209 ATAATATATGATGACTCCAAACACTTATGATTGGTTGCTGCTGAAGCAAAAGTCCCTCAAT 268
Db 371 ATCGGTATGATGATGCTTATCTATCTATGATCTAGTCGGAGCGCGCTCAAAAGTACTCGAA 430
QY 269 CTCATGCTGGAGAAATCCTCCAAATGTTGGGAAGATGTTTTCGCTCTTTGGCCAAGAA 328
Db 431 ATAAGTGGCAATGATCTGTTGGAGGCAATCGGTGGATGTTCTTTGAATTTTGGCTCGAG 490
QY 329 TCTGTTATGATACAAATCTGCTGCTGCTGCTGCTTAATGTACAGAGAAATTTCTACAGAAC 388
Db 491 TCTGCTACGACCAACATCTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
QY 389 CTGATGCTGTCGACGACCACTTCTGCTACCAATCTACCAAGGAATGCTGCTGCTGCTGCTGCT 448
Db 551 TTAGATGCTCTTACGATCATCTTCTGCTCAATCTACCGGGGAATCGGAGCACTTCAATTC 610
QY 449 AGGTGCACTGATGACAGAAAGGCAAGGCAAGGACTCATTTTGCACCTACTACTACAGAGAA 508
Db 611 AGATGCACTACAGAGAGAGTGACGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670
QY 509 GGACTTCAGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 568
Db 671 GGGCTGGAGCACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 730
QY 569 GAATATGACATGAAGGTTATTCAGCAAGAAATGAAGAAATGTGATCATCTACTCAATTTTAA 628
Db 731 GAAATGCAATGCTGAGATCATCAAGAAATGAAGAAAGAGTGGGATCATGTTCAAGTTTGG 790
QY 629 ATTGAAGAAAGACTCAAAAGAGAGGATTTTATGAAGATCTTGAAGATCTTGAAGATTTGAAGAA 588
Db 791 ATCATCGAAGAGGTGGAGACAGCAAGATTTGAAGAGAGGCGGAGCAAGCTT---ACTG 847
QY 689 AATGTATCCAGGAATACGCTAGCCCATATACATCTGCAAGAGCTTTTCTTTTTCAT 748
Db 848 GCTCTCTCAAGGAACCAAAATCAGTCCATCAACACTGTCGAGGATCCTGCCCTTCCAC 907
QY 749 ATAATATTTAGCCGGGACCTAGTGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 808
Db 908 ATCATGTTCAATGCTGTAATTAACGCTAGAGCAAGCAGGTAAGTCCATTTCAAGAGGATTTGA 967
QY 809 CCCCAGCTCCAGCTTGGGAATTTGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 868
Db 968 CCTACATATTCATCTTCACTGCAAGATGAGGAGCTTATTCATATCGTGGGCGCTCAC 1027
QY 869 ATTGATATTTAGTTTCCATGGGATCCTTTCTCACATCAATCTGTTTGTATTTGAGAAGC 928
Db 1028 ATGGAGTTTCACTTCAAGTCTATCTCTCAGTCAATGCAACACATTTACGTCGTGAAGACC 1087
QY 929 AAGGAGGATTTGATGATGTTGGAGAAATAGATGTAGGATGAAGTGAAGTGAAGTGAAGTGAAG 988
Db 1088 AACTCTGTTGTTGAATTTCAACAA-----ACCCCGTAACGGCTCC 1129
QY 989 ATCAGCTGCTTACGCTCAAGGCTCAAAATGATCTACTTACCTGAAGCAGATGATCACTT 1048
Db 1130 ATCCAGCAGCTCAAGCTCAAGGCGGATGCTTACGTAACAGATGCAAGTACTACTA 1189
QY 1049 TTTCTATGTTTCAACAGTGTATGAACTTGGACGATTTTGAAGAGAGAGGCTGTATCTA 1108
Db 1190 TATCTCTGTTCAACACAGCTCATCAATCTGATGAGTGAAGCAAGTGAAGTGAAGTGAAGTGAAG 1249
QY 1109 AGTGACATCTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1168
Db 1250 AGTGATATCTTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309
QY 1169 GAGGAATCAAACTCACCAAGAACTGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
Db 1310 GAGGAGTACAACTAACCCAGAGCTAGAGATCTTCAAGAGCAAGCTACCAAGACCTAC 1369
QY 1229 AGAGCCCTGGAGATGAAGAAAGAAAG 1288
Db 1370 CTTGAGATCGAAGATGAAGAAAG 1429

QY 1289 TCTGTTGCCAATGAGCTGCGGCACAAGCGTCCAGTCCCTGCCAAAAGATATGACAATGTG 1348
Db 1430 TCCGTTGCTAACGAGTTACGTCATCATCGACCTGTTCCGCCCAAAAGTTTGAGTGTG 1489
QY 1349 ACCATCCTTCTTAGTGGCATTTGTTGGCTTCAATGCTTCTTGTAGCAAGCATCTCTGGA 1408
Db 1490 ACTCTGATGTTTCACTGAGGATCTTTGGCTTTGGAGACTTTTGT-----CGTGGCTACTCA 1543
QY 1409 GAAGGAGCATGAAGATCGTCAACCTCTCAACGACCTCTACACAGATTTTACACACTG 1468
Db 1544 CACGAGCTATCAAGATCGTTAGCCTTCTTACAGTGTTTACACCAAGTTTGACGCTCTT 1603
QY 1469 ACTGATTCGCGGAAAACCCATTTCTTATAGTGGAGACTTGTGGTGAAGTATATG 1528
Db 1604 ATGGA-----ATAACCCAGACGTTTACAGGTAGAGACAGTGGGTGATTAATACATG 1657
QY 1529 ACAGTGAAGTGGTTTACCAGGCAATGCAATTCACCATGCAAGTCCATCTGCGACCTGGCC 1588
Db 1658 GCAGTAAGTGGTCTTACCCGTACCATGTGTCATGTAATGATCGCTAAGATGGG 1717
QY 1589 TTGGACATGATGGAATTTGCTGGCCAGGTTCAAGTAGATGGTGAATCTGTTCCAGATAACA 1648
Db 1718 CTTGAATGAAGAACTCTCTGCAATGTCATCATGAGGAGTACCCTATAGTATCACT 1777
QY 1649 ATAGGATACACACTGGAGAGGTAGTTACAGGTGTCATAGGACGCGATGCTCGATAC 1708
Db 1778 ATTGGTGTCTACAGTGGAGAGGTGGTACAGAGTGGTGGGACAACGCGATCGGAGGTAC 1837
QY 1709 TGTCTTTTGGGAATCTGTCACCTCACAACCCCAAGAAACACAGGAGAAAGGA 1768
Db 1838 TGTCTTTTGGTAACTGTCACCTCCAGGACAGAGACTGTTGTTCCAGGCG 1897
QY 1769 AAAATAAATGTCTGTAATATACATACAGATGCTTATGTCCTCCAGAAAATTCAGATCA 1828
Db 1898 AAGATCAATCGCGGATCTCTTACGATGTTTGTATGAGGACCAAGATGCAACCCA 1957
QY 1829 CAATTCCTTGGAGCAGAGGCGGCTGTCATGAGGCGAAAGAAACCAATGCA 1888
Db 1958 ACCTTCCAGTTCGATTTCCGAGGTCTAGTCAACATGAAGGCAAGCTTAACCTGCCCA 2017
QY 1889 GTTGGTCTTATCCAGAAA 1909
Db 2018 TGCTATCTGCTCTCAAGAAA 2038
AF062751 3691 bp mRNA linear INV 25-SEP-1998
Manduca sexta soluble guanylyl cyclase beta-1 subunit mRNA,
complete cds.
AF062751
AF062751.1 GI:3372755
Manduca sexta.
Manduca sexta.
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Sphingioidea; Sphingidae; Sphinginae; Manduca.
REFERENCE
1 (bases 1 to 3691)
AUTHORS
Nighorn, A., Gibson, N.J., Rivers, D.M., Hildebrand, J.G. and
Morton, D.B.
TITLE
The nitric oxide-cGMP pathway may mediate communication between
sensory afferents and projection neurons in the antennal lobe of
Manduca sexta
J. Neurosci. 18 (18), 7244-7255 (1998)
MEDLINE
9840973
PUBMED
9736646
REFERENCE
2 (bases 1 to 3691)
AUTHORS
Nighorn, A., Gibson, N.J., Rivers, D.M., Hildebrand, J.G. and
Morton, D.B.
TITLE
Direct Submission
JOURNAL
Submitted (04-MAY-1998) ARL Division of Neurobiology, University of

Arizona, Gould-Simpson Bldg. Rm 611, Tucson, AZ 85721-0077, USA
FEATURES
Location/Qualifiers
1. .3691
/organism="Manduca sexta"
/db_xref="taxon:7130"
168. .1970
CDS
/note="GC-bl"
/codon_start=1
/product="soluble quanylyl cyclase beta-1 subunit"
/protein_id="AAC61264.1"
/db_xref="GI:3372756"
/translation="MTGFVYALELLVMTFDEETWETIKKKADVAMEGSFLVRIQIE
DEITNLITAAVEVLIQIPADALELFGKTFEFCQDSYDKILQVLGATPRDFLQNL
GLHDLGTLIPKSPSPEDPGLVHLVYSDRPGLEHIVIGIVKTVASKLHNT
EYKVELTKKECDRVOFLITSTTGRVSAIEAIEIETLSLEPKVSPATFCRVERPH
LMPEDRLNIVQAGRTSRLLPRTVPGCKITDVLTVRPHLEMTFANVLAHINTVYVL
KPKPMSYTDPHREIASLRIGQMLYIPETDVVVFQCPVSVTNLDLRLRGICLADI
PLHADRDLVMSFEADYKLTQNLVLELQOTFRELELEKQOTDRLLYSVLPI
VATERHRPVPARYDYDTLLFSGIVGFANYCARNSDRHKAMKIVRLMDLTYAFDV
LTPDKRNPVYKVTGYKMAVSGLPETEVAAKHISLLALDMDLDSQTVTVDGEPV
GITIGHSCEVTVGIGHRMPRYCLFGNTVNLTSRCETTVGPVTINVSBDTVNLMRE
DNHDEQFELTYRGHVTMKGKAPMOTWELTRKH"
BASE COUNT 1105 a 726 c 742 g 1118 t
ORIGIN
Query Match 22.9%; Score 559.6; DB 3; Length 3691;
Best Local Similarity 58.1%; Pred. No. 1.3e-119;
Matches 1060; Conservative 0; Mismatches 739; Indels 27; Gaps 3;
QY 85 CACCATGACGATTTGTGAATCAGCCCTGAGTGTGCTGGTATCCGCAATTTACGGCCC 144
DB 164 CAAATGTACGGGTGTGTGAATCAGCCCTGAGTGTGCTGGTATCCGCAATTTACGG 223
QY 145 CGAGGTGTGGGAGACATCAAAAAGAGGCACAGTGTAGATGAAGAGGACAGTTTCTGT 204
DB 224 AGAAACATGGAGACGATTAAGAAAAGCCGACGTGCTATGGAAGGACAGTTTCTGT 283
QY 205 CAGATTAATATATGATGACTCAAAACTTATGATTTGGTGTGCTGCAAGCAAGTCT 264
DB 284 GAGACAGATTTATGAGATGATGATCACTTATAATTTGATCAGCGCTGCCGTGAGTCT 343
QY 265 CAATCTCAATGTGGAGAAATCTCCAAATGTTTGGGAAGATGTTTCTGCTTTTGCCA 324
DB 344 GCMAATACCCGCTGACGCTATATGAGCTATTTGGAAAACGTTCTTCGAGTTTGCCA 403
QY 325 AGAATCTGTTATGATACAACTTTCGCTGCTGCTGCTTAATGTGAGAAATTTCTACA 384
DB 404 AGATTCGGGTATGATAAATTTCTGCAACTTTCGAGCCACGCGACGAGATTTCTACA 463
QY 385 GAACCTTGATGCTGCGACGACACCTTCTGCTACCATCTACCCAGGAATCGGTGACCTTC 444
DB 464 GAATCTGGACGGCTGACGACATCTCGGGACATTTATATCTCGGATGAGATCCGCCCTC 523
QY 445 CTTTAGTGTCACTGATGACAGAAAAGGCAAGGACACTATTTTGCACCTACTACTCAGAG 504
DB 524 GTTCCGATCCAGCGCTCTGAGATGGCGCCCTTGTGCTCATTTATTTCTCGACAG 583
QY 505 AGAAGGACTTCAGGATTTGTCTTTGGAATCATCAAAACAGTGGCAACAATCCATCG 564
DB 584 GCCTGCTGGACATATGTTATTTGGAATTTGAAGACTGTGGCAGCAACTCCATTA 643
QY 565 CACTGAATATGACATGAGGTTATTCAGCAAGAAATGAAGATGTGATCATCTACTCAAT 624
DB 644 CACCGA--AGTGAAGGTGGAATCTTAAAGACCAAGAGGAATGTGACCAAGTCCAGTT 700
QY 625 TTTAATTTGAAGAAAAGAGTCAAAAGAGAGATTTTATGAGATCTTGACAGATTTC 684
DB 701 CTTGATCACTGAACCTTCCACCACCGGTAGGTGTCTGCACCTGAGATGACAGAGATT-- 758
QY 685 AGAAATGTACCCAGGAATCAGCATACGCCCATATACATTTCTGCAAGCTTTTCCCTTT 744
DB 759 -GAAAGTTGTCTCTTGACCAAAAGTGTGATCCAGGACATTTCTGCCGAGTGTTCATT 817

QY 745 TCATATAAATATTGACCGGACCTAGTGTCTACTCAGTGTGCAATGCTATATACAGAGT 804
DB 818 CCATTTGATGTTTGTGATCGTGAACATCGTCCAAAGCGGACGCTACTGTCTCGTCT 877
QY 805 TCTCCCCAGCTCCAGCCCTGGAAATTTGAGCCCTTCTGTCTGTCTCTCTCCGCTGGTCC 864
DB 878 ACTCCCAAGAGTTTACCAGACAGGCTGTAAAGATTACTGATGCTCTTGATCTACTGTCG 937
QY 865 TCATATTGATATATTAGTTTCCATGGGATCCTTTCTACATCAATACATCTGTTTCTATTGAG 924
DB 938 CCATTTAGAGATGACGTTTTCCTGCTGCTCATATTAACACGGTCTACGTTTGTAA 997
QY 925 AAGCAAGGAAGATTGTTGGATCTGGAGAAATAGAAATGTGAGGATGAATGACTGCTGGAC 984
DB 998 AACCAACCTGAAGAGATGAGTCTACTGATC-----CITCATGA 1036
QY 985 TGAGATCAGCTGCTTACGTCTCAAGGTCAAATGATCTACTTACCTGAAGACAGATAGCA 1044
DB 1037 GGAATTCGCTTCTTCTGCTTGAAGGTCAAATGCTGTACATTCAGAAACAGACGCTGT 1096
QY 1045 ACTTTTCTATCTTACCAGGTGTATGAACCTGGACGATTTGACAAGGAGAGGCTGTA 1104
DB 1097 TGTATTCAGTGTATCTCAAGTGTGACCAATTTGGACGATCTGCTGCCGCGCCCTATG 1156
QY 1105 TCTAAGTGACATCCCTCTGCAATGATGCCACGCGCATCTTGTCTTTTGGGAGACAAT 1164
DB 1157 CATTTGCTGACATTCACCTTTCAGGATCGGACACGAGACCTGTTCTCATCTCCGAGCAAT 1216
QY 1165 TAGAGAGGAATCAAACTCACCACCAAGAACTGGAATCTCTACCTGACAGGCTACAGTCA 1224
DB 1217 TGAAGCTGATTAAGCTGACCCAGAACTGTGAAGTTTGTGCTGACAACTTCAACAAAC 1276
QY 1225 GTTAAAGAGCCCTGGGAAGATGAAGAAAAGAACAGACACATTTGCTGTATTTCTCTTCC 1284
DB 1277 ATTCCGAGAGCTTGAATTTGGAACACAGAGACGACAGCTCTTTTACTCAGTCTGCC 1336
QY 1285 TCCGCTGTGTGCAATGAGCTGCGGACAAAGCGTCCAGTGTGCTGCAAAAGATATGACAA 1344
DB 1337 TATAAGCGTGGCAGCTGAACTTCGTCACGTCGCCGCTACGAGCGCGCTTACGACAC 1396
QY 1345 TGTGACCATCTCTTTAGTGGCATTTGGGCTTCAATGCTTCTGTAGCAAGCATGCA 1404
DB 1397 CGTGACCTTCTCTTACGGGTATTGTAGGGTTCGCCAACTACTGCGGTAGGAATCTGA 1456
QY 1405 TGGAGAAGAGCATGAGATCGTCAACCTCTCAACGACCTCTACACAGATTTTGACAC 1464
DB 1457 CCATAAGGCGCTATGAGATCGTGAATGCTCAATGATCTGTACACCGCTTTGATGT 1516
QY 1465 ACTGACTGATTTCCCGAATAAACCCATTTGTTTATAGGTGGAGACTGTGGTGACAAAT 1524
DB 1517 TCTCACCAGTCTTAAGAGGAATCTTAATGTGTATAAGGTGGAGACTGTGGGTGATAATA 1576
QY 1525 TATGACAGTGTGTTTACCAGAGCCATGATCATTACCATGACGATCCATCTGCCACCT 1584
DB 1577 CATGCCGTATCGGGCTTACCAGATATGAGGTAGTCTATGCTAAACATTTTCGCTTCT 1636
QY 1585 GGCCTTGGACATGATGGAATTTGCTGGCAGGTTCAGGTAGATGTTGATCTGTTCCAGAT 1644
DB 1637 TGCTCTGGACATGATGGACTTGTCCCAACCGTACTGTGGATGGAGAACCACTGGGCAT 1696
QY 1645 AACATAGGGATACACACTGGAGAGGTAGTTACAGGTGTCTAGGACAGCGGATGCTCG 1704
DB 1697 AACCATCGGAATACACTCAGGCGAAGTGTGATCTGTCATCGCTACCGCATGCCAG 1756
QY 1705 ATACTGTCTTTTGGGAATCTCAACCTCAACGCGCAACAGAAACACAGAGGAGAAA 1764
DB 1757 CTACTGTCTCTTCCGCAATACAGTGAATTTGACCAAGCGCTGTGAGACTTCTGGAGTACC 1816
QY 1765 GGGAAAAATAATGCTCTGTAATATACATACAGATGCTTGTGTCTCCAGAAAAATTCAGA 1824
DB 1817 TGGCCACATTAACGTCAGGAGGACACTTACAAATTAACCTTATTCGCGGAGACACCA 1876
QY 1825 TCCCAATTCCTACTTGGAGACAGAGGCCCTAGTGTCCATGAAGGGGCAAAAAAGAACCA 1884

```

Db      1877 CGACAGTTGAGCTGACATCGTGGTCACGTACCATGAGGGGAGGCGAGAACCCCAT 1936
QY      1885 GCAAGTTGGTTCTTCATCCAGAAAA 1910
Db      1937 GCAGACCTGGTTCTTCACACGAGGA 1962

RESULT 12
AC114761
LOCUS   Homo sapiens BAC clone RP11-354G23 from 4, complete sequence.
DEFINITION
AC114761
VERSION AC114761.3 GI:21307587
KEYWORDS HTG.
SOURCE  human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 70709)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
MEDLINE
9847074
PUBMED
2 (bases 1 to 70709)
Buatsi,D., Bielicki,L. and Creason,K.
The sequence of Homo sapiens BAC clone RP11-354G23
Unpublished (2001)
3 (bases 1 to 70709)
Waterston,R.H.
Direct Submission
Submitted (11-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 70709)
Waterston,R.H.
Direct Submission
Submitted (30-APR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 70709)
Waterston,R.H.
Direct Submission
Submitted (01-JUN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 70709)
Waterston,R.
Direct Submission
Submitted (12-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jun 1, 2002 this sequence version replaced gi:20340523.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0354G23
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-588K22, 2000 bp overlap. Actual start of this clone is at base position 88662 of RP11-588K22; actual end is at base position 70709 of RP11-354G23.

Polymorphisms have been identified between AC021433 and AC114761.

Data from AC021433 was used to finish this clone, AC114761.

FEATURES

FEATURES	Location/Qualifiers	Source
repeat_region	1..70709	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"	
repeat_region	/chromosome="4"	
repeat_region	/map="4"	
repeat_region	/clone="RP11-354G23"	
repeat_region	/clone_lib="RPCI-11"	
repeat_region	2..101	
repeat_region	/rpt_family="MaLR"	
repeat_region	155..590	
repeat_region	/rpt_family="MaLR"	
repeat_region	1228..1592	
repeat_region	/rpt_family="MaLR"	
repeat_region	1593..1726	
repeat_region	/rpt_family="MIR"	
repeat_region	1871..2030	
repeat_region	/rpt_family="MIR"	
repeat_region	2905..2925	
repeat_region	/rpt_family="(CAA)n"	
repeat_region	3372..3615	
repeat_region	/rpt_family="MIR"	
repeat_region	5097..5399	
repeat_region	/rpt_family="Alu"	
repeat_region	6270..7028	
repeat_region	/rpt_family="L1"	
misc_feature	8582..9872	
repeat_region	/note="CpG island (8GC-71.1, o/e=0.70, #CpGs=119)"	
repeat_region	12310..12409	
repeat_region	/rpt_family="MER1_type"	
repeat_region	12422..12462	
repeat_region	/rpt_family="MER1_type"	
repeat_region	12867..12978	
repeat_region	/rpt_family="L1"	
repeat_region	13369..13537	
repeat_region	/rpt_family="MIR"	
repeat_region	15390..15574	
repeat_region	/rpt_family="L2"	
repeat_region	15920..16147	
repeat_region	/rpt_family="L1"	
repeat_region	16537..16741	
repeat_region	/rpt_family="MIR"	
repeat_region	18011..18038	
repeat_region	/rpt_family="AT-rich"	
repeat_region	18645..18713	
repeat_region	/rpt_family="CT-rich"	
repeat_region	18757..18900	

repeat_region	/rpt_family="MIR"	19796. .19907	1983	AGGTTCCAGTTTTCTCCTAACACACGTGCCAAGCCAGGAGCAGTTCTTCCCTATGGATACA	2042
repeat_region	/rpt_family="CRI"	20476. .20717	Db	56082	AGGTTCCAGTTTTCTCCTAACACACGTGCCAAGCCAGGAGCAGTTCTTCCCTATGGATACA
repeat_region	/rpt_family="L2"	22037. .22066	Qy	2043	GATTTTCTTTTGTCCCTGTGTCATTAACCCCAAGACTTTCTTCTAGATATATCTCTCATAT
repeat_region	/rpt_family="(TTTG)n"	22070. .22357	Db	56142	GATTTTCTTTTGTCCCTGTGTCATTAACCCCAAGACTTTCTTCTAGATATATCTCTCATAT
repeat_region	/rpt_family="Alu"	22891. .22950	Qy	2103	CCGTTATTCACACCTTAGCTCTGCTTTCTATTACTTTTTAGCTTTTACGCTTATATATCTCTCA
repeat_region	/rpt_family="L1"	23147. .23412	Db	56202	CCGTTATTCACACCTTAGCTCTGCTTTCTATTACTTTTTAGCTTTTACGCTTATATATCTCTCA
repeat_region	/rpt_family="L2"	23413. .23624	Qy	2163	GTGGGCTTTTGTGATGGATGATGAGCTTCATGCTGCTTAAATCTACTACAAGCAT
repeat_region	/rpt_family="Alu"	23625. .24290	Db	56262	GTGGGCTTTTGTGATGGATGATGAGCTTCATGCTGCTTAAATCTACTACAAGCAT
repeat_region	/rpt_family="L2"	24309. .24331	Qy	2223	ACCTAACATGTTGATGTCGAAGTAGTAGGCCACCAATAAATATTTGTTGAATTTAGTTAA
repeat_region	/rpt_family="AT-rich"	24931. .25231	Db	56322	GTGGGCTTTTGTGATGTCGAAGTAGTAGGCCACCAATAAATATTTGTTGAATTTAGTTAA
repeat_region	/rpt_family="Alu"	25653. .26056	Qy	2283	ATGAACTGAACAGTGTGTTGGCCATGTGTATATTTATATCATGTTTACCAAACTGTTTA
repeat_region	/rpt_family="L1"	26057. .26323	Db	56382	ATGAACTGAACAGTGTGTTGGCCATGTGTATATTTATATCATGTTTACCAAACTGTTTA
repeat_region	/rpt_family="Alu"	26324. .26432	Qy	2343	GTGTTCCACATATATGATATATATTTTAAATGACTATATATGTAATAAAGTTTATATCA
repeat_region	/rpt_family="L1"	26557. .26906	Db	56442	GTGTTCCACATATATGATATATATTTTAAATGACTATATATGTAATAAAGTTTATATCA
repeat_region	/rpt_family="MER2_type"	28704. .28787	Qy	2403	TGTTGGTGTATATCATTTATAGAAATCATTTTCTAAAGAGT
repeat_region	/rpt_family="MER1_type"	28788. .28926	Db	56502	TGTTGGTGTATATCATTTATAGAAATCATTTTCTAAAGAGT
repeat_region	/rpt_family="Alu"	28927. .28966			
repeat_region	/rpt_family="(CATA)n"	28968. .29116			
repeat_region	/rpt_family="MER1_type"	29512. .29965			
repeat_region	/rpt_family="L2"	31294. .31344			
repeat_region	/rpt_family="AT-rich"	31712. .32026			
repeat_region	/rpt_family="Alu"	32194. .32590			
repeat_region	/rpt_family="MaLR"	33530. .33559			
repeat_region	/rpt_family="AT-rich"	35905. .36095			
repeat_region	/rpt_family="Alu"	36096. .36410			
repeat_region	/rpt_family="Alu"	36411. .36544			
repeat_region	/rpt_family="Alu"	37678. .38283			
repeat_region	/rpt_family="MaLR"	38324. .38464			
repeat_region	/rpt_family="MIR"	40010. .40030			
repeat_region	/rpt_family="(TTTTTG)n"	40569. .40613			
repeat_region	/rpt_family="Mariner"	40758. .41052			
repeat_region	/rpt_family="Alu"	41089. .41398			
Query Match		21.3%; Score 521; DB 9; Length 70709;			
Best Local Similarity		100.0%; Pred. No. 1.1e-110;			
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1923	AGGAACAAAGCAGGATGACTGAATCTTGATTTATGGGTGAAGAGGACTACAGACT	1982		
Db	56022	AGGAACAAAGCAGGATGACTGAATCTTGATTTATGGGTGAAGAGGACTACAGACT	56081		

Bozulavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castie, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Hesford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlungu, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE JOURNAL

COMMENT

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6778505.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5027

Center clone name: 688_K_17

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 170986 bases at least Q40

Consensus quality: 174287 bases at least Q30

Consensus quality: 175619 bases at least Q20

Insert size: 174000; agarose-fp

Insert size: 176456; sum-of-contigs

Quality coverage: 5.1 in Q20 bases; agarose-fp

Quality coverage: 5.0 in Q20 bases; sum-of-contigs.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 2292: contig of 2292 bp in length
* 2293 2392: gap of 100 bp
* 2393 4948: contig of 2556 bp in length
* 4949 5048: gap of 100 bp
* 5049 8358: contig of 3310 bp in length
* 8359 8458: gap of 100 bp
* 8459 12531: contig of 4073 bp in length
* 12532 12631: gap of 100 bp
* 12632 17861: contig of 5230 bp in length
* 17862 17961: gap of 100 bp
* 17962 28411: contig of 10450 bp in length
* 28412 28511: gap of 100 bp
* 28512 40270: contig of 11759 bp in length
* 40271 40370: gap of 100 bp
* 40371 56792: contig of 16422 bp in length
* 56793 56892: gap of 100 bp
* 56893 80742: contig of 23850 bp in length
* 80743 80842: gap of 100 bp
* 80843 111092: contig of 30250 bp in length

* 111093 111192: gap of 100 bp
* 111193 144467: contig of 33275 bp in length
* 144468 144567: gap of 100 bp
* 144568 177556: contig of 32989 bp in length.

FEATURES

source

1..177556

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone-lib="RPCI-11 Human Male BAC"

1..2292

/note="assembly_fragment"

2393..4948

/note="assembly_fragment"

5049..8358

/note="assembly_fragment"

8459..12531

/note="assembly_fragment"

12632..17861

/note="assembly_fragment"

17962..28411

/note="assembly_fragment"

clone_end:17

vector_side:right

28512..40270

/note="assembly_fragment"

40371..56792

/note="assembly_fragment"

56893..80742

/note="assembly_fragment"

clone_end:SP6

vector_side:left

80843..111092

/note="assembly_fragment"

111193..144467

/note="assembly_fragment"

144568..177556

/note="assembly_fragment"

BASE COUNT 55171 a 32590 c 32371 g 56323 t 1101 others

ORIGIN

Query Match 21.3%; Score 521; DB 2; Length 177556;
Best Local Similarity 100.0%; Pred. No. 1.1e-110;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1923 AGGAACAACAGCAGATGATGACTGAATCTTGGATTATGGGTGAAGAGGATACAGACT 1982
DB 43237 AGGAACAACAGCAGATGATGACTGAATCTTGGATTATGGGTGAAGAGGATACAGACT 43296
QY 1983 AGGTTCCAGTTTTCCTTAACACGTGCCAAGCCAGGACAGCTTCTCCCTATGGATACA 2042
DB 43297 AGGTTCCAGTTTTCCTTAACACGTGCCAAGCCAGGACAGCTTCTCCCTATGGATACA 43356
QY 2043 GATTTTCTTTTCTCCTTGTCCATTTACCCCAAGACATTTCTTCTAGATATATCTCACTAT 2102
DB 43357 GATTTTCTTTTCTCCTTGTCCATTTACCCCAAGACATTTCTTCTAGATATATCTCACTAT 43416
QY 2103 CGGTTATTCACCTTTAGCTCTCTTCTTCTTATTTAGGCTTTAGGTATATATCTATAA 2162
DB 43417 CGGTTATTCACCTTTAGCTCTCTTCTTCTTATTTAGGCTTTAGGTATATATCTATAA 43476
QY 2163 GTTTGGCTTTTGTGATGATGTGAGCTTCATGTCTTAAATCTACTACAAGCAT 2222
DB 43477 GTTTGGCTTTTGTGATGATGTGAGCTTCATGTCTTAAATCTACTACAAGCAT 43536
QY 2223 ACCTAACATGGTGATCTGCAAGTAGTACGCCACCCCAATAAATATTTGTTGAATTTAGTTAA 2282
DB 43537 ACCTAACATGGTGATCTGCAAGTAGTACGCCACCCCAATAAATATTTGTTGAATTTAGTTAA 43596
QY 2283 ATGAACATCAACAGTGTGGCCATCTGTATATATATATATATATATATATATATATATAT 2342

34545 35422: contig of 878 bp in length
35423 gap of unknown length
36294: contig of 872 bp in length
36295 gap of unknown length
37161: contig of 867 bp in length
37162 gap of unknown length
38046: contig of 885 bp in length
38047 gap of unknown length
38943: contig of 897 bp in length
38944 gap of unknown length
39821: contig of 878 bp in length
39822 gap of unknown length
40724: contig of 903 bp in length
40725 gap of unknown length
41524: contig of 800 bp in length
41525 gap of unknown length
42428: contig of 904 bp in length
42429 gap of unknown length
43297: contig of 869 bp in length
43298 gap of unknown length
44181: contig of 884 bp in length
44182 gap of unknown length
45034: contig of 853 bp in length
45035 gap of unknown length
45922: contig of 888 bp in length
45923 gap of unknown length
46818: contig of 896 bp in length
46819 gap of unknown length
47715: contig of 897 bp in length
47716 gap of unknown length
48609: contig of 869 bp in length
48609 gap of unknown length
49477: contig of 869 bp in length
49478 gap of unknown length
50381: contig of 904 bp in length
50382 gap of unknown length
51251: contig of 870 bp in length
51252 gap of unknown length
52145: contig of 894 bp in length
52146 gap of unknown length
53042: contig of 897 bp in length
53043 gap of unknown length
53940: contig of 898 bp in length
53941 gap of unknown length
54805: contig of 865 bp in length
54806 gap of unknown length
55676: contig of 871 bp in length
55677 gap of unknown length
56569: contig of 893 bp in length
56570 gap of unknown length
57455: contig of 886 bp in length
57456 gap of unknown length
58312: contig of 857 bp in length
58313 gap of unknown length
59194: contig of 882 bp in length
59195 gap of unknown length
60086: contig of 892 bp in length
60087 gap of unknown length
60940: contig of 854 bp in length
60941 gap of unknown length
61835: contig of 895 bp in length
61836 gap of unknown length
62731: contig of 896 bp in length
62732 gap of unknown length
63619: contig of 888 bp in length
63620 gap of unknown length
64445: contig of 826 bp in length

Query Match 18.5%; Score 452.2; DB 2; Length 143255;
Best Local Similarity 95.7%; Pred. NO. 1.2e-94;
Matches 463; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
1960 TGGGGTGAAGAGGAGTACAGTAGTCCAGTCTTCTCTTAACAGGTGCCAAGCCAGG 2019

Db 20378 TNGAAGAGGAGTACAGNACTAGGGTTCAGTCTTACTCCTTAACAGCTGCCAAGCCAGG 20437
QY 2020 AGCAGTCTTCCCTATGGATACAGATTTCTTTTGTCCCTTGCTCCATTATCCCAAGACTTTT 2079
Db 20438 AGCAGTCTTCCCTATGGATACAGATTTCTTTTGTCCCTTGCTCCATTATCCCAAGACTTT 20497
QY 2080 CTTCTAGATATATCTCTCACTATCCGTTATTCACACTTACGCTCTGCTTTCTTACTTTT 2139
Db 20498 CTTCTAGATATATCTCTCACTATCCGTTATTCACACTTACGCTCTGCTTTCTTACTTTT 20557
QY 2140 TAGGCTTTAGTATATATCTCTCACTTAAAGTTTGGCTTTTGAATGGATGATGCTGATG 2199
Db 20558 TAGGCTTTAGTATATATCTCTCACTTAAAGTTTGGCTTTTGAATGGATGATGCTGATG 20617
QY 2200 TCTTAAATCTACTACAAGCATTACCTAACATGGTGATCTGCAAGTAGTAGGCACCAAT 2259
Db 20618 TCTTAAATCTACTACAAGCATTACCTAACATGGTGATCTGCAAGTAGTAGGCACCAAT 20677
QY 2260 AAATATTTGTTGAATTTAGTTAAATGAAGTGAACAGTGTGGGCATATATATTTAT 2319
Db 20678 AAATATTTGTTGAATTTAGTTAAATGAAGTGAACAGTGTGGGCATATATATTTAT 20737
QY 2320 ATCATGTTTACCAATCTGTTTAGTGTCCACATATATGATATATATTTAAAGT 2379
Db 20738 ATCATGTTTACCAATCTGTTTAGTGTCCACATATATGATATATTTAAAGT 20797
QY 2380 ATAATGTAATAAAGTTTATATCATGTTGGTGATATATATATAGAAATCATTTCTAAAG 2439
Db 20798 ATAATGTAATAAAGTTTATATCATGTTGGTGATATATATATAGAAATCATTTCTAAAG 20857
QY 2440 GAGT 2443
Db 20858 GAGT 20861

RESULT 15
DMU27123
LOCUS
DEFINITION Drosophila melanogaster soluble guanylyl cyclase beta subunit (dgcbl), complete cds.
ACCESSION U27123.1 GI:861202
KEYWORDS
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 2840)
AUTHORS Shah,S. and Hyde,D.R.
TITLE Two Drosophila genes that encode the alpha and beta subunits of the brain soluble guanylyl cyclase
J. Biol. Chem. 270 (25), 15368-15376 (1995)
MEDLINE 95318108
PUBMED 7797526
REFERENCE 2 (bases 1 to 2840)
AUTHORS Shah,S. and Hyde,D.R.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-1995) David R. Hyde, Biological Sciences, Univ. of Notre Dame, Galvin Life Science Building, Notre Dame, IN 46556, USA

FEATURES
source
1. .2840
/organism="Drosophila melanogaster"
/strain="Oregon-R"
/db_xref="taxon:7227"
/chromosome="3"
/map="100B"
1. .2840
/gene="dgcbl"
300. .2576
/gene="dgcbl"

gene
CDS

```
/codon_start=1
/product="soluble guanylyl cyclase beta subunit"
/protein_id="AA87941.1"
/db_xref="GI:861203"
/translation="MYGFVNALELLVLKHFGEIWEIKKKAMVNEGOFVRLQYD
DEITYNLGNAVEILNIPADDILELFGKFFFCQSDGYDKILQVLGATPRDFLQND
ALHDHGLTLPGRAPSFRCERKDELLHYISERFGLHEHIVIGVAVASAKLHGVVEV
EIDIVRKEPIDEAREKARARENQOLLEDAVATTTGSAVYVLAAPSSDAERNNNH
GSGSNNGMANNVTNNVNNNDQOIASSETDPSIALSTCPAODSFDCDQKQK
LRLKKNKDDIRYDVOFLIREINVAASKOVDAKKDEVPDDMEFLCEAPLISPATFC
KVPFHLMDROMKIQVAKAVSRVPRVAEENCLEVEEARPHQLNFENLISHI
NIYVLTQRCGAMSSHEQFRLKGMWYIPETDRILLFCYFVSNLDDLTKKGLYI
SDVPLHDAADLVLLSEKFAEYKLTNLEMLFDKLOQFFRDLESKECTDRLLYSVL
KPSVANELRHQRPVPRKRVDSVTFMFSIGVFGQYCAANTDPDGMKIVKMLNELTYV
FDALITSKRLNLYKVTYGDYKMYAVSLPDHCEHAKCMARVALDMDMAMKVKMGK
NPQVITIGHSGEWTGVIIGNRVPYCLFGNTVNLTSRTETGVPGRINVSRETYRL
CMAINODDSPHLEVRGPIVIMKGRPTPMDCHFRLTRATSSILGGTSSGSGGGGNGSLD
SPLLQATPAPAGGPRFP"
BASE COUNT      730 a   753 c   782 g   573 t   2 others
ORIGIN
Query Match      15.2%; Score 371.4; DB 3; Length 2840;
Best Local Similarity 58.1%; Pred. No. 9.7e-76;
Matches 694; Conservative 0; Mismatches 471; Indels 30; Gaps 1;
QY 705 CACGCATCAGCCCATATACATTCGCAAGCTTTTCCCTTTTCATATAATTTGACGGG 764
DB 1276 CACTCATATACGCCGCCACTTTTCGAAGGTGTTCCCTTCCACCTGATGTTGATAGGC 1335
QY 765 ACCTAGTGGTCACTAGTGGCAATGCTATATACAGAGTTCCTCCCGACCTCCAGCTG 824
DB 1336 AATGAAATCGTTCAGCGCCGCAAGCTGTGCTCGAGTAATTCGCAAGTTGCCGAGG 1395
QY 825 GGAATTCAGCCCTTCTGTCTCTTCGCTGCTGCTCTCATATGATATTAGTTTCC 884
DB 1396 AGAAGTTCCTTCCCTAATAGAGGTGGTGAGGCCATCGCGCGACCTGCACCTCACTTCG 1455
QY 885 ATGGGATCTTCTCACATCAATACACTGTTTCTTATGAGAAGCAAGGAAGATTGTTGG 944
DB 1456 AGAATATCTCTCCACATAAACCACATCTATCTCTGACAGACAGCTCAGGGGGCCATGA 1515
QY 945 ATGCGGAGAAATAGATGTGAGGATCACTGACTGGGACTGAGTCACTGCTTACGTC 1004
DB 1516 GC-----AGCGGACAGAGCAGCGGTTTCTGAGAC 1545
QY 1005 TCAAGGTCAAATGATCTACTTACCTGAAGCAGATAGACATACCTTTTCTATGTTACACAA 1064
DB 1546 TGAAGGGCAATGATGTACATTCGGGAACGATCGGATACCTGTTCAATGCTATCCCA 1605
QY 1065 GTGTCATGAACCTGGAGGATTTGACAAGGAGAGGGCTGTATCTAAGTGACATCCCTCTGC 1124
DB 1606 GTGTCATGAATCTGGATGATCTAACCAAGAGGAGCTTACATCTCGGATGTGCCGCTGC 1665
QY 1125 ATGATGCCAGCGCGATCTGTTCTTTTGGAGAACAAATTTAGAGAGGATACAAACTCA 1184
DB 1666 ACAGCGTGCAGGGATTTGGTCTCTGCTCCGAGAAATTCGAGCGGAGTACAAGCTCA 1725
QY 1185 CCCAAGAACTGGAATCCTCACTGACAGGCTACAGCTCACGTTTAAAGAGCCCTGGAAGATG 1244
DB 1726 CCAAGAACCTTGAATGCTCACCGACAAGCTGACGAGACATTCGCGATCTCGAGAGCG 1785
QY 1245 AAAAGAAAAGACAGACACATTCCTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1304
DB 1786 AGAAGCAGAAAGACCGCAGGCTGCTCTATTCGGTGTCTGCAAAAGTCGGTGGCCAAATGAGT 1845
QY 1305 TCGCGGACAGGCTCCAGTCCCTGCAAAAGATATGACATGTGACCATCTCTCTTTAGTG 1364
DB 1846 TAGACATACGCGCTCTGTGCGCCCAACACGCTACGCTCCGCTGAGCTGTGATGTTCTCCG 1905
QY 1365 GCATTGTGGGCTTCAATGCTTCTGTAGCAAGCATGCTATCGGAGAGGAGCGCATGAAGA 1424
DB 1906 GCATGTGGGCTTTGGCCAATCTGTGCGGCCAACACAGATCCCGACGGGGCAATGAAA 1965
```

Search completed: July 1, 2003, 16:02:36
Job time : 4194.4 secs

```
QY 1425 TCCTCAACCTCTCTCAACGACCTCTACACACAGATTTGACACACTGACTGATTCCCGGAAAA 1484
DB 1966 TCCTGAAAAATGCTTAATAGCTCTACACGGCTCTTCGATGCTCTGACCGACTCCAAGAAA 2025
QY 1485 ACCCATTTGTTTATAGGTGGAGACACTGTTGGTCACAAGATATATGACAGTGAAGTGTTC 1544
DB 2026 ACCTGAACGTTGTAACAGGTTGAACACATTTGGCGATAGTACATGCGCGTTTCCGGACTGC 2085
QY 1545 CAGAGCCATCATTCACCTACAGGATCCATCTGCCACCTGGCCTTGGACATGATGAAA 1604
DB 2086 CCATCACTCGGAGGATCATGCCAAGTCATGCGACGAGTGGCCCTCGATATGATGGACA 2145
QY 1605 TTCTGGCCAGGTTCAAGTAGATGTTGTAATCTCTTCAGATAAACAATAGGATACACACTG 1664
DB 2146 TGCCCAAGAACGTCAAAATGGGATCCAATCCAGTGCATAATCACCATAGGATCCATTCGG 2205
QY 1665 GAGAGTAGTTACAGGTGTCATAGGACAGCGGATGCTCGATACTGCTTTTGGGAATA 1724
DB 2206 GCGAAGTGGTGAAGTGGATGATCGCAACCGAGTGCACGCTACTGCTCTGTCGGAATA 2265
QY 1725 CTGTCACCTCAACGCCGACCAACACAGAAACACAGGAGGAAAAAATAAATGTGCTG 1784
DB 2266 CTGTCATCTCACACGCCGACCGAGACCACTGGCTTCGCGGCGGATCAACGTCAGCG 2325
QY 1785 AATATACATACAGATGCTTTATGCTCCAGAAAAATTCAGATCCACAATTCACCTTGGAGC 1844
DB 2326 AAGAAACCTATCGACTCTCTGTATGGCGATCAACGAGGATGATTCGTTCCACTTGGAGT 2385
QY 1845 ACAGAGGCCAGTGTCCATGAAGGCAAAAAAACAACCAATGCAAGTTTGGTTCT 1899
DB 2386 ACCGCGACCCGTTATCATGAAGGCAAAACCGACGCCCATGGACTGTTGTTCT 2440
```